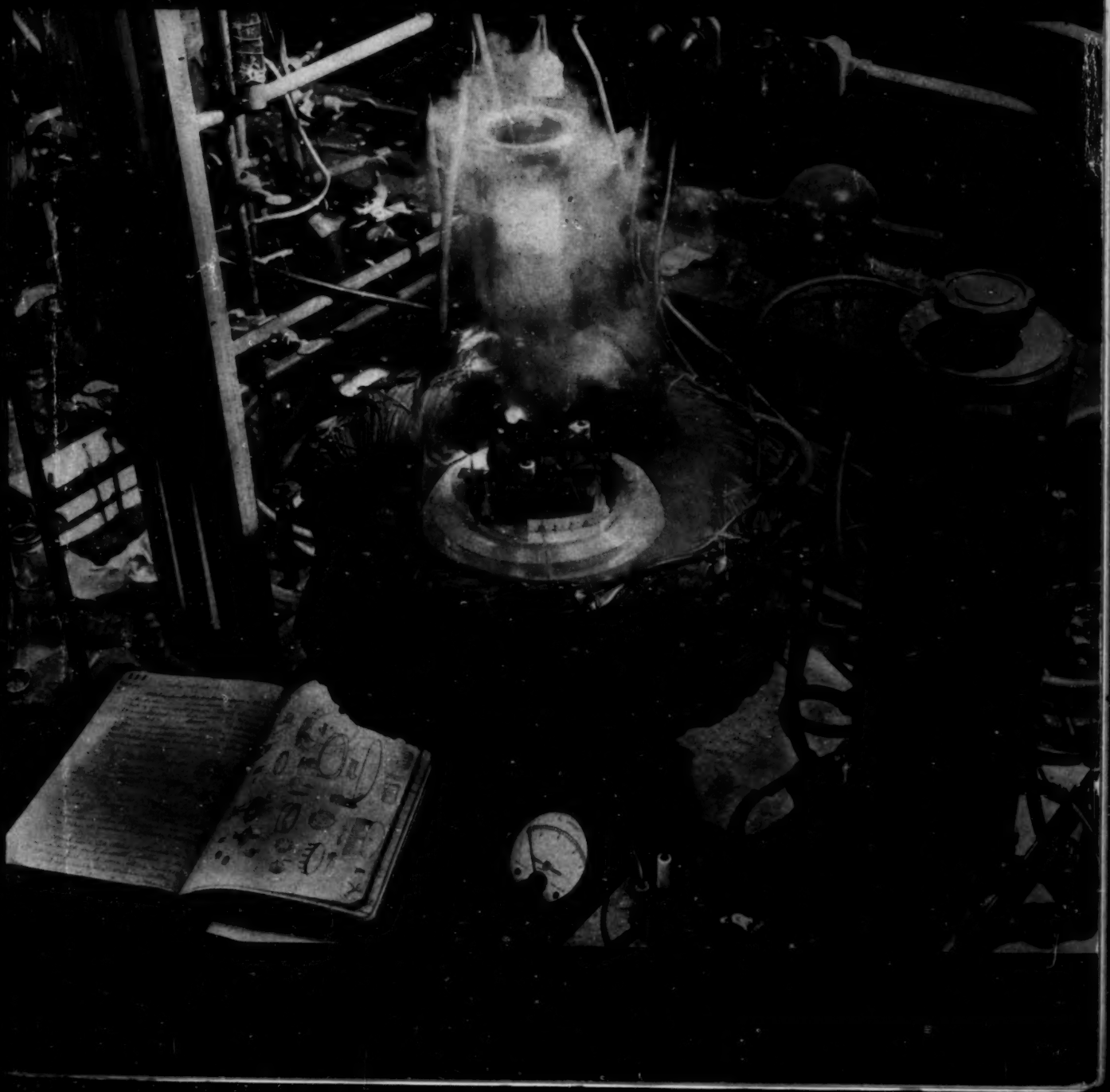
JOURNAL OF APPLIED PHYSICS



JOURNAL
OF
APPLIED PHYSICS

Vol. 19, No. 1 JANUARY, 1948

In This Issue

		Page
Graphite Lubrication	ROBERT H. SAVAGE	1
Conical Dipole of Wide Angle	P. D. P. SMITH	11
Theory of Slots in Rectangular Wave Guides	A. F. STEVENSON	24
Scattering of Electromagnetic Radiation by a Thin Circular Ring in a Circular Wave Guide	PAULA FEUER AND EDWARD S. AKELEY	39
Resonant Cavity Linear Accelerator	A. B. CULLEN, JR. AND J. H. GREIG	47
Temperature Coefficient of Electrical Resistivity for Crystalline Selenium Containing Various Percentages of Bromine	WAYNE E. BLACKBURN	51
Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers	W. C. ELMORE	55
General Divergence Formula	H. J. RIBLET AND C. B. BARKER	63
Theory of the Study of Transfers within a Multi-Compartment System Using Isotopic Tracers	C. W. SHEPPARD	70
On the Stresses in a Plate Containing Two Circular Holes	CHIH-BING LING	77
Alignment Chart Giving the Polarization Correction of Equi-Inclination Weissenberg Photographic Intensities	W. L. BOND	82
Simple Procedure for the Making of Alignment Charts	W. L. BOND	83
Characteristics of Helical Antennas Radiating in the Axial Mode	JOHN D. KRAUS AND J. CLAUDE WILLIAMSON	87
On the Structure of Cellulose Acetate Molecules in Acetone Solution	SEYMOUR J. SINGER AND HERMAN MARK	97
Effect of Ground Constants on the Characteristic Values of the Normal Modes in Non-Standard Propagation of Microwaves	C. L. PEKERIS	102
Errata: The Electrostatic Field of a Point Charge inside a Cylinder, in Connection with Wave-Guide Theory	C. J. BOUWKAMP AND N. G. DE BRUIJN	105
Study on Anaeroid Capsules	EUGENE W. PIKE AND NORMAN E. GIBBS	106
Letters to the Editor:		
Remarks on Slow Waves in Cylindrical Guides	A. A. OLINER	109
Specimen Contamination in Electron Microscopes	JOHN H. L. WATSON	110
Use of a Cavity Oscillator for a Kerr Electro-Optical Light Shutter	A. C. LAPSLEY, L. B. SNODDY, AND J. W. BEAMS	111
Proceedings of the Electron Microscope Society of America		118

In Every Issue

Calendar of Meetings	108
New Books	112
New Booklets	116
Here and There	116
Index to Advertisers	xxxii

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND
JOHN VON NEUMANN

Publication of the
American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

CLEVELAND NORCROSS, *Assistant Director*

RUTH F. BRYANS, *Publication Manager*

THEODORE VORBURGER, *Business Manager*

JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

GEORGE R. HARRISON, *Chairman*

WALLACE WATERFALL, *Secretary*

GEORGE B. PEGRAM, *Treasurer*

E. U. CONDON
A. J DEMPSTER
W. F. FAIR, JR.
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
PAUL E. KLOPSTEG
A. H. PFUND
JOHN T. TATE
M. W. WHITE
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members and associate members of the American Institute of Physics	$6.00	$7.00
To all others	8.00	9.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 18, 1947—$132.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1948, by the American Institute of Physics.

Physics Lib
Fagan

Journal
of
Applied Physics

Volume 19, Number 1 January, 1948

Graphite Lubrication

ROBERT H. SAVAGE
General Electric Research Laboratory, Schenectady, New York
(Received July 14, 1947)

The slipperiness of graphite is not inherent in the crystal structure alone but depends upon adsorption films, especially water, which cover the carbon atoms and provide surfaces of low cohesion. In vacuum graphite rods or brushes seize upon a moving base (copper or graphite) and wear away as a fine dust. This wear and accompanying high friction are instantly stopped by some condensible vapors at low pressures (H_2O, C_6H_6, NH_3, etc.) which apparently effect the coverage of the graphite by a reversible adsorption governed by the pressure and surface temperature. Under conditions of optimum coverage (zero wear) the graphite friction is still appreciable but may be accounted for as the heat of cohesion of water monolayers attached to the carbon atoms and undergoing successive makes-and-breaks. From the measured frictional energy and an assigned heat of cohesion based upon surface energy data, the area of molecular contact is estimated to be 4×10^{-6} cm² under a real pressure 2×10^{4} kg/cm².

The wear dust of unlubricated graphite shows a unique adsorptive power at room temperature, greater for hydrogen than that of activated charcoal by five orders of magnitude. This and other characteristics may throw new light upon the mechanism of friction in general and upon the structure of stroked graphite surfaces in particular.

INTRODUCTION

FOR the past 20 years the accepted picture of graphite lubrication has retained as its basis the notion that the slipperiness is an intrinsic property of the crystal, resulting from the saturated character or low cohesion of the cleavage faces. This conception seemingly originated in the interpretation of the lattice structure deduced from x-ray data. According to W. L. Bragg:[1] "Two neighbors in any one layer (of graphite) are drawn together more closely than two carbon atoms in diamond. But the great distance between layer and layer is naturally associated with weakness in their mutual attraction. It is just this peculiarity that makes graphite so perfect a lubricant. One layer can slide freely over another." This conception has been only slightly modified by the later work of Jenkins and Holm. In 1934, Jenkins[2] showed by electron diffraction that the polishing of graphite surfaces produces orientation of the crystals with the main cleavage plane (001) in the plane of the specimen. This would appear to have been implicit in the lubrication process according to the Bragg conception, where the layers slide freely over one another. However Jenkins' original proof with its elaborations added a new emphasis. Holm[3] further developed the orientation idea and ascribed the occasional wear of graphite surfaces to *tilting* of the crystals such that, as razor edges, they shaved the mating surfaces. He suggested that the graphite scales which form the sliding surface are held to the base by a thin film of water "acting as a glue,"

[1] W. L. Bragg, *Introduction to Crystal Analysis* (G. Bell and Son, Ltd., London, 1928), p. 64.

[2] R. O. Jenkins, Phil. Mag. 17, 457 (1934).
[3] R. Holm, *Electrical Contacts* (Hugo, Gerber, Forlog, Stockholm, 1946), p. 193.

and that if the water is lost as in dry air the scales may tilt so as to produce great wear.

We see that in all of the foregoing the supposedly saturated character of the graphite faces is evidently regarded as the fundamental basis for the lubrication, and that the crystal orientation, presumably essential for maintenance of an interface having low cohesion, received the emphasis when it became necessary to explain why graphite sometimes fails to show low friction.

The present paper is concerned with a new viewpoint on the lubrication, a preliminary account of which has already been published.[4] It is based upon experimental methods aimed at the control of adsorption films upon the carbon. These films are too thin to have been detected by diffraction. In the new work it is found that the slipperiness of graphite depends upon these films. When they are absent the graphite seizes and rapidly disintegrates into a black dust of exceptional sorptive power. The properties of this dust provide important clues regarding the nature of the wearing process and of the lubrication mechanism.

APPARATUS

At the start of this work it was necessary to find bearing materials for the apparatus which would operate in vacuum. The main bearing used at first consisted of a steel doughnut of V-section mating to a graphite race which was polished to a specular finish. This bearing was satisfactory in air, but wore excessively in vacuum and ultimately seized.

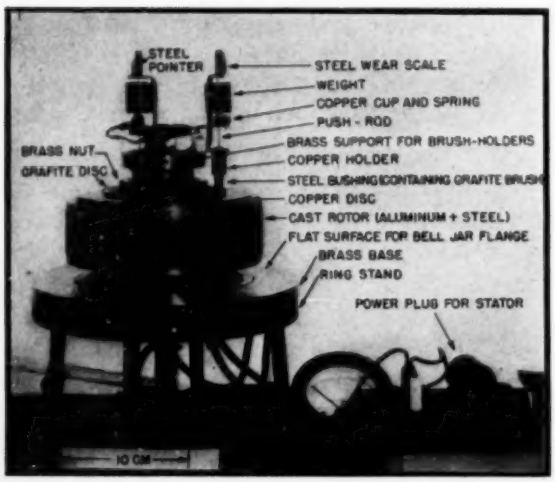


FIG. 1. Sliding contact apparatus (without bell-jar and stator).

After abandoning the graphite bearing, it was found that steel ball-bearings which had been specially cleaned and then coated lightly with Apiezon grease-L would operate temporarily both in vacuum and in air. These were used in the preliminary work, although, as we are to see later, a better lubricant for the steel proved to be powdered graphite coated with a trace of Apiezon grease.

The apparatus is shown in Fig. 1. It consists of a small induction motor, the cast squirrel-cage rotor revolving about its vertical axis within a glass bell-jar and carrying the collector-disk which comprised the moving member of the contact. The disk was copper, of 15-cm diameter, and revolved at about 1750 r.p.m. (1370 cm/sec.). Graphite rods comprising the stationary part of the contact were mounted in brush-holders normal to the plane of rotation, and bore against the disk face under forces which were varied from 0.01 to 1 kg but which were usually held constant at 0.09 kg. The rods were small in section, 1.8 to 6.4 mm^2, so that the linear wear was easily observed during operation and so that the rubbing faces were confined to narrow tracks. The brush systems were mounted on arms which under frictional force deflected concentrically against a steel spring, thus indicating the kinetic friction.

Although at first the sliding contact was one of graphite brushes on copper, the work was extended to cover other disk materials with many carbon compositions; and the conclusions which will be drawn are not necessarily limited to brush contacts but are applied to graphite lubrication broadly.

THE WEAR OF GRAPHITE IN VACUUM

Initially in vacuum the friction of the graphite rods against copper was low and the wear-rate substantially zero. Owing to the inconsistency of this result with the previously noted graphite bearing failure it was supposed that the surfaces were not clean, and attempts were made to drive off possible contaminants by passing high transient currents through the graphite to the spinning copper disk. A fivefold rise in friction

[4] R. H. Savage, Gen. Elec. Rev. **48**, 13 (October 1945).

resulted and the graphite began to wear away as a fine dust. The high rate persisted after removal of the current until the rods were worn out. Later it was noted that the copper surface was not roughened microscopically by the current transients, from which it appeared that the high wear was the result of true seizure. The question of surface contaminants was then studied in detail.

Search indicated several sources of contamination which were contributing to what has proved to be a masking of the intrinsic frictional behavior of graphite, and which is a critical factor in any fundamental experiments on graphite lubrication. These sources were found to be both organic and inorganic in origin. They were finally eliminated (1) by baking-out the apparatus in vacuum (except the bearings) to 400°C, (2) by firing the graphite in vacuum, and (3) by cleaning the copper disk so that its surfaces were hydrophilic and remained so when under vacuum.

The copper-cleaning was particularly important. It was found for example that a disk, which had been polished with fine abrasives and a buffing wheel, retained occluded "grease" so that even though cleaned until apparently hydrophilic it became hydrophobic merely upon standing in vacuum. Because of this absorbing tendency, which was found to affect the lubrication critically, and also because the usual polishing powders (e.g. MgO) are known to taint metal surfaces, at least from a molecular viewpoint, the use of polishing powders or abrasives was excluded. The disk was finely lathe-turned without lubricants (0.0079-cm pitch), and was cleaned by cathodic action with heavy gassing in a solution of sodium carbonate, followed by water rinsing. This left the surface hydrophilic and free from appreciable oxide layer, and it remained so in vacuum.

When after these precautions the graphite rods were operated against the spinning copper disk in vacuum they wore down reproducibly with extreme speed. Typical carbons showed wear-rates of 0.1 mm^3/second. This is the rate of wear-loss by volume, and is equivalent to 16 microns/second linear wear for the rods having 6.4-mm^2 cross section.

THE WEAR-DUST

The wear-product was found to be mainly *black* carbon. The specific surface of this dust was found in some instances to exceed 10^4 cm^2 per gram of carbon. It was found, accidentally, that the fresh dust was extremely active while still under vacuum so that it cleaned up hydrogen at low pressures even though it was not chilled. This unexpected property was first noted in the following way:

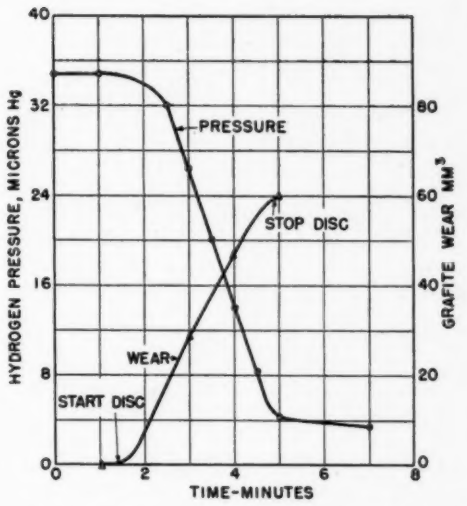


FIG. 2. Relation of hydrogen clean-up to graphite wear.

With the bell-jar exhausted, hydrogen was admitted to a pressure of 38 microns mercury and the system was then closed. The pressure remained constant with the disk stationary. It was intended (in another connection) to dissociate this hydrogen into atoms with a hot filament[5] near the disk but it was found upon spinning the latter that during the brush dusting the hydrogen disappeared at a rate much faster than was expected from the temperature and other conditions of the filament, the rate falling to normal shortly after stopping the disk.

The filament was then left out of the experiment, and the clean-up of hydrogen as the direct result of brush-dusting in this gas at low pressure was observed in detail. The following experiment (Fig. 2) is typical. Upon starting the disk the brush-dusting began at once (within ten seconds). The hydrogen pressure up to this point had been constant (35 microns) but now began a steady fall, as observed with a sensitive Pirani gauge in a bridge circuit. The rate of clean-up reached 0.9-mm^3 hydrogen gas per second (NTP). The run was continued until 60 mm^3 of brush carbon

[5] I. Langmuir, J. Am. Chem. Soc. **34**, 1310 (1912).

TABLE I.

Carbon	Temperature of carbon °C	Gas	Pressure of gas microns	Adsorption (gas vol. at RTP) mm³/gram	Source of data
Charcoal	−195	H_2	6	111	Claude*
Charcoal	room (15)	nitrogen	3.23	0.173	Rowe**
Charcoal	room	H_2	3	0.017	Rowe**-Titoff*
Charcoal	room	H_2	3	0.008	Claude* (extrapolated)
Graphite dust	30	H_2	4	1680	Savage

* S. Dushman, "High vacuum," Gen. Elec. Rev. (1922), Chap. IV.
** H. Rowe, Phil. Mag. 1, 109 (1926).

had been worn off, which required 3.5 minutes. The disk was then stopped quickly, by reversal of the stator field. Simultaneously the clean-up rate decreased rapidly, approaching zero within one minute after standstill.

In the first experiments of this kind a glass trap in the ceiling of the bell-jar was chilled with liquid nitrogen to condense traces of water vapor (which had been found to be the only condensible vapor present in the apparatus at appreciable pressures). It was noted that a little of the carbon dust collected on this trap although most was thrown to the sides of the jar. The experiments were therefore repeated with the trap warm to make sure that the clean-up was *not* due to the small amount of *chilled* carbon. The gas disappearance again occurred at a high rate proving that the chilling of a small amount of the carbon had not been an important factor.

After exhausting the apparatus containing the dust to 0.01-micron Hg, fresh amounts of hydrogen were introduced. These were not cleaned up in the slightest degree showing that the initial hydrogen was irreversibly held, and that the carbon was already saturated with it.

It was found that nitrogen and methane also were cleaned up by fresh dust from wearing carbons.

THE WEARING PROCESS

From experiments of the foregoing kind, we should be able to learn much that is fundamental about graphite lubrication and wear. If the dusting is due to a scaling-away of the graphite crystals, we should expect the particle size of the brush to be reflected in the character of its dust, and to find, for example, a predominance of relatively large crystals in the dust from brushes made from natural graphite. But we do not find this. Brushes made from gray natural graphite flakes (of high optical reflectivity) produced black carbon dust, similar in sorptive activity to that collected from brushes made from fine-particle carbons (graphitized coke and lamp-black).

We are thus led to the view that the wearing process involves a steady subdivision of graphite crystals to minute size including a continual tearing apart of the individual scales *at right angles to the slip-planes*. The scission of intraplanar bonds involved in this is rather remarkable when it is remembered that these bonds are probably stronger than those in the diamond as indicated by their shorter C−C distance (1.41A). The exposed valence forces easily account for the observed activity toward hydrogen at room temperature.

SURFACE PROPERTIES OF THE DUST

The active surface indicated by the hydrogen probably represents only a fraction of the total. Nevertheless this activity provides a key to the basic wearing process; and this fact warrants further consideration of the clean-up rate and an estimate of the specific surface.

In a representative experiment the rate of hydrogen clean-up $v=0.91$ mm³/second with a volume wear-rate for the two brushes $2w=0.286$ mm³/second. The corresponding wear-rate by weight $2w_g=0.543\times10^{-3}$ gram/second. The hydrogen adsorption is thus

$$v/2w_g=0.91/0.543\times10^{-3}=1680 \text{ mm}^3/\text{gram}. \quad (1)$$

Assuming an area per hydrogen molecule $a=6\times10^{-16}$ cm², the specific surface of the dust active to hydrogen is

$$S_h=2.69\times10^{16}va/2w_g=2.7\times10^4 \text{ cm}^2/\text{gram}. \quad (2)$$

The sorptive power of this graphite in comparison with that of activated charcoal is shown in Table I. The graphite adsorption is seen to be 10^5 times that of the charcoal. We should note that the clean-up by the graphite dust occurred

without the flow of current through the graphite rods, and that it was repeated by preparing the fresh dust in vacuum and then admitting the hydrogen, showing that the frictional heating was not an essential factor.

There are few published data on the adsorption of gases at low pressures and moderate temperatures such as are required for direct comparison. The data shown for the adsorption of hydrogen by charcoal at room temperature were obtained by extrapolating Claude's higher pressure data through four decades on a log-log graph, which gave the adsorption value 0.008 mm^3/gram; and also by multiplying Rowe's low pressure data for nitrogen by the H_2/N_2 ratio indicated by Titoff's data at 0°C and 100 mm, which gave the value 0.017 mm^3/gram. Both of these methods are crude. But the results check within a factor of about 2 and are compatible with the data for nitrogen at 15°C and for hydrogen at −195°C. They thus provide at least a rough basis for comparison with the brush dust. It is seen that the wear-product in vacuum is a "super-activated" black carbon derived from graphite by purely mechanical means.

The foregoing has been concerned with the wearing process. We have now to consider the *lubrication* of graphite.

THE VAPOR LUBRICATION OF GRAPHITE

When the graphite brushes on the copper disk were exposed to water vapor the high friction coefficient ($\mu = 80$ percent) which had persisted in vacuum showed an immediate drop. At a pressure of 3 mm the wear-rate had fallen to practically zero ($< 10^{-4}$ mm^3/second), and the friction which had been fluctuating violently became low ($\mu = 18$ percent) and constant. This effect was found to be quickly reversible, an equilibrium being reached between wear-rate and water vapor pressure as in Fig. 3. At pressures above 3 mm the brushes could be operated indefinitely with negligible wear.

When hydrogen was substituted for water vapor, the graphite rods dusted away at wear-rates approaching those observed in vacuum. No lubricant effect with this gas was noted even at a pressure of 600-mm Hg. When this enormous difference in the effects of gases upon graphite lubrication was found,[6] a series of trials of other vapors and gases was made.

Nitrogen and carbon monoxide showed no lubricant effect, at least in the pressure range up to 600 mm, the graphite wearing in these gases as in vacuum. Oxygen did show a lubricant effect, similar to that of water but at pressures higher by a factor of about 100; i.e., a pressure of 200 to 400 mm was required to lower the wear-rate to the zero range noted with water at about 3 mm. On the other hand, the easily condensible vapors which were tried produced complete lubrication (i.e., low wear-rate) at pressures below 5 mm, including ammonia, acetone, benzene, ethanol, diethyl ether, hexamethyldisiloxane,* n-hexane. In general the efficiency of the gas or vapor in the lubrication seems to increase with boiling point (or ease of condensation), although there is indicated an additional specific factor related to composition, as shown by the difference between oxygen (bp, −183) and nitrogen (bp, −196).

MECHANISM OF LUBRICATION

The high wear of the graphite in vacuum and the powerful effects of certain condensible vapors

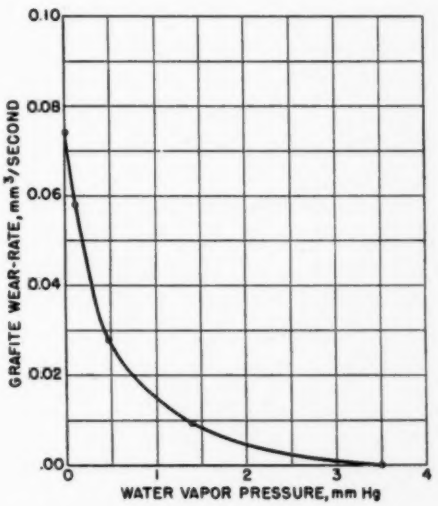


FIG. 3. Graphite wear-rate *vs.* water vapor pressure.

[6] C. Van Brunt and R. H. Savage, Gen. Elec. Rev. **47**, 16 (October 1944).

* This vapor belongs in a special category. Unlike the others but characteristic of many silicones it is easily oxidized at the sliding interface to form highly abrasive products (siliceous residues) as will be described in a later article. It is included here because when oxygen is excluded it is found to lubricate the graphite as well as do the other vapors such as water.

in inducing low friction indicated that the lubricating characteristic is not intrinsic in the graphite crystal alone but depends upon adsorption films. Guided by the condensation-evaporation theory of adsorption the writer has suggested[6] that the vapor molecules tend to cover the exposed graphite surface by forming transient monolayers, thus lowering the free surface energy. Purely as an adsorption concept this was old, but as a mechanism of lubrication it appeared to be new. The essential effects involved, as played by the gas molecules, are described in the usual way by their rate of bombardment of the surface

$$n_b = 3.535 \times 10^{22} p/(MT)^{\frac{1}{2}} \text{ molecules/cm}^2 \text{ second}, \quad (3)$$

where p is the gas pressure in mm mercury, M is the molecular weight, and T is the absolute gas temperature; and by their rate of evaporation from the surface

$$n_e = Ae^{-Q/RT_s}, \quad (4)$$

where A is a constant, Q corresponds to an "activation" energy for the evaporation of water from the surface, and T_s is the surface temperature.

THE LIFE OF THE WATER MOLECULES UPON CARBON

These rate considerations enable us to form a more fundamental picture of the lubrication process by developing the kinetic aspects. In the case of water, for example, we may take the minimum lubricating pressure of 3 mm (Fig. 3) as that which produces the completed monolayer over the carbon. From Eq. (3) the corresponding rate of bombardment is 1.4×10^{21} molecules/cm^2 second. Assuming that the condensation coefficient $\alpha=1$, the average life[7] of these molecules on the carbon is

$$t = N_s/n_b, \quad (5)$$

where N_s is the number of molecules covering unit area. Assuming $N_s = 10^{15}$ molecules/cm^2

$$t = 10^{15}/1.4\times10^{21} = 0.7\times10^{-6} \text{ second}.$$

The life of this assumed film on carbon is seemingly too short to be measured directly. However, a qualitative check has been attempted. In this a glass nozzle was mounted in the bell-jar near the face of the graphite brush and directed at the track on an area which would reach the brush face in 10^{-3} second during rotation. A thin-walled heated capillary tube was arranged outside of the jar but in series with the nozzle so that water could be evaporated under vacuum directly onto the track, as a blast having a mass equal to 10 molecular layers per revolution of the disk (1.8×10^4 molecular layers per minute).

With this arrangement the bell was exhausted and the trap chilled with liquid nitrogen. On spinning the disk, the graphite brush showed a high wear-rate and high friction. The water blast was then applied for one minute. The high wear and friction persisted without change. Since the friction is extremely sensitive to transient frictional changes, this experiment showed that the life of the molecules on the track was less than 10^{-3} second.

The effect of the cold trap was obviously to condense the evaporating vapor before it could strike the track near the brush a second time. When a sudden flood of water vapor (more than enough to saturate the trap momentarily) was let into the vacuum by other means, a transient down-swing of the friction arm occurred, indicating momentary lubrication. The effect disappeared within a few seconds as the trap cleaned up the water.

From the foregoing experiments it is easily inferred that the gases of very low boiling points cannot be expected to form continuous films for coverage of the graphite at the low pressures which are found effective in the case of the vapors. Hydrogen, for example, probably has an adsorption life so short that a pressure of many atmospheres (or an under-cooling of the carbon) would be required for lubrication. This brings us to what may seem a curious anomaly:

Hydrogen at very low pressures is irreversibly adsorbed by the clean graphite which is produced as dust at the interface, and immediately saturates it; and yet hydrogen at nearly atmospheric pressure fails to lubricate the graphite, apparently because it re-evaporates too rapidly ($t<10^{-6}$ second).

SUGGESTED EXPLANATION OF THE HYDROGEN ANOMALY

This behavior of hydrogen is easily explained if we assume that the seizure which is associated

[7] I. Langmuir, J. Chem. Soc. p. 521 (1940).

with the high wear occurs at both the edge atoms and the face atoms of the graphite crystal, and that the adsorption characteristics are completely different for the two types, an assumption in accord with the established idea of "active areas." On the edge atoms we assume that the hydrogen sticks irreversibly as suggested by the clean-up effect at low pressures; on the face atoms that it condenses and evaporates with speeds more nearly approaching those of a reflection, thus failing to cover the graphite.

On the faces the extent of the adsorption should increase with pressure and decrease with temperature according to Eqs. (3) and (4). Since the effective surface temperature is not constant the equilibrium required for lubrication is appropriately expressed as an isostere (not as an isotherm); i.e., as a plot of minimum lubricating pressure *vs.* surface temperature for constant amount (area) of adsorbed gas. Since isosteres are fundamentally similar to vapor pressure curves for liquids we have to consider the exponential increase in evaporation rate with increase of surface temperature as a key factor in the lubrication, as will be seen in the following section.

GRAPHITE ON GRAPHITE

The described effects of wear and lubrication were observed using graphite rods against the spinning copper disk, and occurred without the flow of current through the rods. The effects were found to be modified by current, but not profoundly (as will be described in a later article). The graphite rods which were tried included pressed natural graphite, coke-bonded natural graphite, electrographitized coke, and electrographitized coke-bonded black carbon. Because their essential constituent is graphitic carbon, we call all of these graphite even though the differences in apparent physical properties are considerable. With the exception of *impure* natural graphite all of these showed high wear in vacuum and responded to vapor lubrication. So it began to appear that the lubrication picture as developed for the graphite-copper combination might be applied more broadly. For this a knowledge of the films which formed on the brush track was essential.

Fortunately an analysis of the character and composition of the films on the copper had been completed[8] by Van Brunt several years before the present study was started so that the concepts which had been formed were available to build upon. The initial analysis (1937)—which had been performed entirely by microchemical methods—isolated 3.2 milligrams of brush track from a copper ring, and revealed essentially a two-layer structure of cuprous oxide and carbon, the latter covering the oxide so as to form a continuous graphite sliding surface of average thickness 3.3×10^{-5} mm. This was the first isolation of a graphite film by chemical means. One of its important implications was that the sliding occurs between graphite surfaces rather than between graphite and copper.

When a graphite disk** was substituted for copper in our friction experiments the dusting occurred in vacuum as before, which supports the view that the seizure occurs between graphite surfaces. However when exposed to water vapor the C−C system showed unexpected differences in comparison with the C−Cu system, related primarily to the sliding speed.

The peripheral speed of the disk was given as 1370 cm/second at 1750 r.p.m., and since the graphite rods were operated against its face upon different tracks (shorter radii) in the different experiments the sliding speed was varied in steps down to about 600 cm/second. It was possible to lower this range still further in a rough way to about 10 r.p.m. (34 cm/second) by controlling the stator field intensity.

It was found that the low pressure of water vapor (3 mm), which had been completely effective for the C−Cu system, lubricated the C−C system only in the low speed range, i.e., below about 200 cm/second. In this range low friction and zero wear-rates were consistently noted. But at higher speed, higher water pressures were required to maintain lubrication; for example, at 800 cm/second the minimum pressure was 6 to 7 mm.

A peculiarity of this speed effect was the abrupt transition from low friction to actual dusting which occurred repeatedly when the water pressure was held at 4 mm and the speed increased from the low range to maximum. Instability de-

[8] C. Van Brunt, unpublished reports (1937); also Gen. Elec. Rev. **47**, 28 (August 1944).

** Acheson plate, turned to a clean face with dry tools.

TABLE II. Examples of the relation between friction and wear-rate for graphite against copper.

Atmosphere	Wear-rate (mm^3/second)	Friction (μ)	Wear-rate/friction ($\times 8$)
Vacuum	0.1	0.8	1.0
Water vapor	<0.0001	0.18	<0.004

veloped at some critical intermediate speed, within a few seconds, and the subsequent high friction and dusting could be stopped only by decreasing the speed to the low range or by increasing the pressure to that required for the high speed. These transitions were repeated many times.

With the copper disk no marked speed effect had been noted up to 1370 cm/second. It was, therefore, supposed that the speed effect for graphite was due to the higher surface temperatures which may be expected to result from its lower conductivity. On the supposition that the temperature gradient through the surface layers of the graphite disk increased with speed, it is apparent that the water layer will become depleted by increased evaporation at the given pressure; and also that when friction begins to increase during the initial stages of this depletion, because of increasing frequency of seizure, the gradient will further increase even at constant sliding speed so that the system becomes unstable and the abrupt transition to dusting follows. From this it would seem that the copper (as disk material) has the important function of quenching the carbon film so that the heat transients developed continuously at the sites of frictional contact can raise the effective surface temperatures only rather slowly with increasing speed. Much more work is needed in order that the separate contributions of thermal conductivity, load and speed can be understood in relation to these surface temperatures and to the minimum lubricating pressure. Our attempts to determine these temperatures by the thermocouple method (as developed by F. P. Bowden and others) have been confused by the immediate formation of the graphite film on the copper disk which results in a graphite-graphite contact.

MECHANISM OF FRICTION

The relation between the wear-rate and the friction of graphite is not a simple one. Their ratio is about 1 (in arbitrary units) during high wear in vacuum, but this ratio falls below 0.004 when the wear-rate becomes low in a water vapor atmosphere (Table II). Thus the friction during low wear is still much higher than might be anticipated. We have, therefore, tried to account for the heat of friction.

It was natural to start with the assumption that the frictional heat developed during the dusting stage could be accounted for as the energy required to break the large number of C−C bonds in the seizure process. It was surprising to find that even for the case of severe dusting, the power required to open these bonds at the rate indicated by such experiments as those on hydrogen clean-up accounted for less than 0.06 percent of the total frictional power. This estimate was based on a C−C bond energy[9] of 58.6 kcal./mole (which we assumed to have a mechanical significance equivalent to its thermal one).

On the other hand, if it is assumed that the hydrogen clean-up accounts for only a fraction (e.g., .01 percent) of the total bonds opened, the majority being the weaker ones between slip-planes which presumably do not clean up hydrogen, the frictional power might be accounted for. But the total specific surface of the dust must then exceed 10^8 cm^2/gram, which is beyond possibility.

So we are led to conclude that the frictional power during dusting is not accounted for by the bond scissions which occur in the wearing process. Furthermore, since for the case of lubricated graphite (water vapor atmosphere) the friction is considerable when the wear is negligible (Table II), we may conclude that bond scission (or a tearing apart of the graphite) is not the source of *normal* (i.e., lubricated) friction. For example, the average frictional power of a typical graphite rod in our apparatus is roughly 1.8 watts when the wear is virtually zero (under 3-mm water vapor pressure, 1000 cm/second sliding speed, 0.092-kg normal force). This represents a considerable rate of heat loss for the condition of negligible seizure or damage to the contact surfaces. We are thus led to speculate upon the cause of the friction.

[9] L. Pauling, *Nature of Chemical Bond* (Cornell University Press, Ithaca, New York, 1939), p. 53.

We may assume that the basic frictional process for graphite lubricated in the vapor atmosphere is one involving cohesion and adhesion, and some of the most useful suggestions are those of Tomlinson[10] in one of the few papers attempting to account for heat of friction. According to his theory the frictional process (for metals) includes the approach, the contact, and the recession of one atom above another; and these involve, respectively, (1) the attraction fields, (2) the interaction of valence bonds, with support of the load by the repulsion fields, and (3) the severing of the atomic bond. In this make-and-break process it is the "plucking action" of one atom on the other, during the break, which evolves heat. The writer considers that the irreversibility of the proposed work cycle by which this is accomplished may be questionable, but regardless of this the conclusion reached by Tomlinson is extremely suggestive; *viz*, that the quantity of energy dissipated during each make-and-break is independent of the velocity of the relative motion.

There is another way to account for the heat of friction which makes it unnecessary to go into the actual vibrational behavior of the disturbed surface atoms but which still retains the molecular viewpoint as well as a consistency with the principles of adsorption. We have already shown evidence that the sliding normally occurs between carbon atoms covered with monolayers of water. When two such surfaces approach and make contact there should be evolved the heat of cohesion. The quantity of this energy per pair of molecules will be independent of speed. Unlike the mechanism of Tomlinson this heat appears during the make, and his basic assumption of heat evolution during the break is omitted. Since the time of relaxation[11] is small compared with the time of an individual molecular contact (10^{-11} second at the sliding speed 1000 cm/second) the heat is easily absorbed by the system. Then, under the conditions of a lowered energy state due to the cohesion, work must be done to separate the molecules. This is the frictional work and serves to draw the slider atoms to the next contact site, where the process repeats.

[10] G. A. Tomlinson, Phil. Mag. **7**, 907 (1929).
[11] S. J. Gregg, *Adsorption of Gases by Solids* (Methuen and Company, Ltd. London 1934), p. 59.

This frictional work is assumed to be that required to break (or to pull apart) a column of water equal in cross section to the real area of brush contact so as to produce two separated faces of twice this area. This assumption is based upon the well-known Dupre expression for the work of adhesion, which for two like surfaces (water) becomes numerically equal to twice the surface tension, γ, in ergs per square centimeter. If we make the additional assumption that γ for the water monolayers on the carbon is approximately equal to that of a pure water surface, the heat of cohesion for an accumulated number of makes-and-breaks totalling one square centimeter of molecular contact may be estimated as $W_c = 140$ ergs/cm^2 $= 1.4\times10^{-5}$ joule/cm^2. Thus the relatively large frictional power of the non-wearing graphite may be accounted for simply through a series of surface cohesions and interruptions which occur without surface damage.

This mechanism of friction enables us to estimate the real area of cohesive contact, a quantity which is of fundamental importance to the physics and engineering of sliding contacts.

AREA OF COHESIVE CONTACT

The calculation of this area is made for the normal case of the "non-wearing" or lubricated graphite. It is based upon the supposition that the frictional power, P, of the brush (1.8 watts in the example given) is due to the rate of expenditure of the work of cohesion, W_c.

The rate of cohesion of the water-covered graphite surfaces may be written as

$$r_c = P/W_c = 1.8/1.4\times10^{-5} = 1.3\times10^5 \text{ cm}^2/\text{second}. \quad (6)$$

It is evident that a rod making complete or ideal molecular contact with the disk would show an enormously greater rate of cohesion R_c than the actual rod, and that the ratio of the real or cohesive area of contact a_o to the apparent or projected area A is equal to the ratio of the cohesion rates; i.e.,

$$a_o = Ar_c/R_c. \quad (7)$$

The ideal rate R_c is simply the product of the number n_f of sites on the projected face of the rod, the rate l at which rows of sites on the track are crossed by those on the rod, and S the area

per site. At the speed 1000 cm/second in the example

$$R_c = n_f l S$$
$$= (0.062/10^{-15})[1000/(10^{-15})^{\frac{1}{2}}](10^{-15})$$
$$= 2\times 10^{9} \text{ cm}^2/\text{second}. \quad (8)$$

The area of cohesive contact is thus

$$a_c = 0.062\times 1.3\times 10^{5}/2\times 10^{9}$$
$$= 4\times 10^{-6} \text{ cm}^2.$$

The minuteness of this quantity deserves special note. The real pressure on this area, under the normal force 0.092 kg, is 23000 kg/cm². This is below the probable upper limit for the intrinsic compressive strength of carbon.[12] Thus the calculated area lies within reasonable limits. Increase in force should increase this area by elastic deformation or by compressive failure but not by plastic flow. The constancy of μ for varying load, observed experimentally, indicates that this increase does occur and suggests that the net area is directly proportional to the load.

Under the extreme pressure indicated, a strong anchoring of the crystals to the base would seem essential to confer stability upon the graphite film. The unsaturated forces of the crystals should be sufficient for this. Assuming that the crystals tend to lie flat (W. L. Bragg and R. O. Jenkins) the carbon atoms are in the position of maximum surface packing and within the crystal should show an intrinsic hardness against normal penetration approaching that of diamond.[13] For this reason the graphite crystal structure is important in the lubrication, a conclusion which should be emphasized along with the additional effect of surface coverage by the vapors, which has been our main subject.

It is hard to see now how a film composed of graphite crystals oriented with their slip-planes exactly flat on the base (as referred to in the Introduction) could withstand the vertical force component without penetration between adjacent platelets and resultant scaling-off of the film. We are uncertain about the true structure of the surface layers. Tentatively it may be assumed that the graphite platelets do not lie flat as has hitherto been believed, but overlap in one direction (according to the direction of stroke) to form a *shingled* layer. Experiments supporting this view will be described in a later article.

ON GRAPHITE FRICTION IN RELATION TO METALLIC FRICTION

From a molecular viewpoint the graphite sliding surface may now be looked upon as a series of overlapping networks of carbon atoms, bonded individually by the powerful intraplanar forces which characterize graphite and covered by monolayers of condensed vapor such as water which present a surface of low cohesive energy. This is analogous to a boundary oil film (covering a metal) which consists of overlapping backbones (chains) of carbon atoms each covered with saturated groups such as bonded hydrogens.

For both graphite and metal systems the friction becomes high when the saturated groups are removed. When this occurs we suppose that the basic cause of the friction (i.e., cohesion) remains the same although increasing greatly in magnitude. In the seizure process which then occurs, primary valence bonds develop so that the identity of the interface is lost at the minute sections in cohesive contact. When this seizure involves a concentrated area of appreciable magnitude the metal system, as is well known, undergoes distortion deep into the masses of the parts. In contrast the porous and weak cleavage structure of the graphite permits local fragmentation, which is the dusting process, and this proceeds without the distortion of contour of the moving base which is characteristic of metallic seizure.

The most remarkable difference between graphite and metallic lubrication seems to hinge on the fact that the graphite obtains its saturated surface from the normal atmosphere (water vapor and oxygen) by a process which evidently limits the action to a monolayer. This natural coverage of the graphite goes on effectively even under severe thermal loading imposed by electric power, with actual current densities at the interface (to be discussed later) which can now be estimated as exceeding 10^6 amperes/cm². Without this unique ability of graphite to conduct with high efficiency across its own lubricant film the modern process of current collection or commutation would appear impossible.

[12] P. W. Bridgman, J. App. Phys. **12**, 461 (1941).
[13] P. C. L. Thorne and E. R. Roberts, *Inorganic Chemistry* (Ephraim), Gurney and Jackson, London, 1943, p. 22.

The Conical Dipole of Wide Angle

P. D. P. SMITH
Montreal, Quebec, Canada
(Received February 13, 1947)

The following report deals with a method of calculating the admittance of dipoles that consist of complete cones of wide angle. The theory makes use of the orthogonal properties of Legendre's functions and of their derivatives in order to make the outside field's tangential component vanish over the spherical end surfaces of the dipoles, and to make the inside and outside fields fit at the boundary sphere. This is a surface of the same radius as the dipole, extending from the upper to the lower end surface.

This leads to a set of Eqs. (10) for the infinite set of coefficients b_k, and an Eq. (3) that expresses the effective load admittance y_t in terms of the coefficients b_k. If a sufficient number of these coefficients is considered, there should be no appreciable discontinuities in field at the boundary sphere, and the tangential field should vanish on all the surfaces of the dipole.

On the assumption that only the first two outside waves, and the internal T.E.M. and first higher order inside wave are important for dipoles whose radius is less than a half-wave-length, an approximate formula for y_t is found, and the results are shown in Figs. 4 and 6. It is unnecessary to evaluate the solutions of Legendre's equation $L_n(\theta)$, because the integrals that involve them are expressed in terms of $n(\mu_1)$ and $dn(\mu_1)/d\mu_1$, n being the root of $L_n(A)=0$, and $\mu_1=\cos A$.

The solutions are transformed to the center of the dipole, and plotted in Fig. 6, while the relative terminating admittances, $K\cdot y_t$, are plotted on an impedance diagram in Fig. 7.

THE DIPOLE PROBLEM

ONE of the most important problems of communication theory is that of the antenna. There are many different phases to this problem, roughly falling into two general classes, those dealing with the spatial distribution of the radiation, and those dealing with the impedance that the radiator presents to its feeding transmission line.

Radiation may be emitted from horns, loops, slots in wave-guides, and dipoles. These elements may be used alone, or combined in arrays. In this report only the problem of a single dipole's impedance is considered, the dipole being formed of the two sides of a metallic cone, with ends which are sections of a sphere.

The different shapes of dipole which have been analyzed may be arranged in a sort of order of successive deformations, which is cyclic in that the last shape returns to the first. These shapes are: (1) The thin wire, analyzed by assuming a current distribution, and summing the effects of current elements upon the field. (2) The cylindrical antenna, investigated by L. V. King;[1] E. Hallen; Ronold King and F. G. Blake; Ronold King and C. W. Harrison.[2] (3) Thin conical antennae, investigated by Schelkunoff.[3] (4) The conical antenna of wide angle, which is the subject of this paper. (5) The spherical antenna, described by Stratton and Chu.[4] (6) The spheroidal antenna, also described by Stratton and Chu;[5] and by Page and Adams.[6]

The successive variations of form are shown in Fig. 1. The fourth of these forms covers the range from the narrow cone of Schelkunoff to the spherical dipole of Stratton and Chu, and is used in practice for broad-band, short-wave dipoles.

The dipole problem can be handled either by obtaining an integral equation for the current distribution, or by fitting characteristic solutions of Maxwell's equations to certain boundary conditions. The first method has been largely used for the second form,[2,7] the second method for the last four forms.[3–5]

Schelkunoff[3] has approached the dipole problem by showing that the fields at the apex of a conical dipole of radius 1 are similar to those near the apex of a conical transmission line terminated, at a radius 1, by a certain admit-

[1] L. V. King, Phil. Trans. **A** 381 (1937).
[2] R. W. King and C. W. Harrison, J. App. Phys. **15**, 170 (1944).
[3] Schelkunoff, Proc. I.R.E. **29**, 493 (1941).
[4] J. A. Stratton and L. J. Chu, J. App. Phys. **12**, 236 (1941).
[5] Stratton and Chu, J. App. Phys. **12**, 241 (1941).
[6] Page and Adams, Phys. Rev. **47**, 819 (1938).
[7] D. Middleton and R. King, J. App. Phys. **17**, 273 (1946).

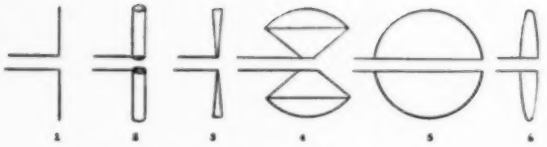


FIG. 1. Forms of dipole that have been analyzed mathematically.

tance y_t. This very useful analogy depends upon the form taken by the principal, or T.E.M., waves guided by the cones, and holds within a radius where the higher modes can be neglected.

Since the properties of transmission lines are well known, the problem is that of finding the proper form of the admittance y_t. This is done by making the appropriate solutions of Maxwell's equations continuous at the "boundary sphere" (which is shown in Fig. 2), and making the tangential electric fields vanish at the dipole surfaces. The formula found for y_t is only an approximation, and the boundary conditions are not exactly met, because the expansions for the fields are limited to a finite number of terms.

The transmission line analogy depends upon taking $V(r)=\int E_\theta \cdot r d\theta$ and $I(r)=\int H_\phi \cdot r \cdot \sin A \cdot d\phi$ as the analogs of the line e.m.f. and current. If this were done, $V(r)$ would depend only on the T.E.M. waves, but $I(r)$ would contain additional terms of higher order. The analogy holds near the apex where the higher terms are negligible, and here $I(r)$ and $V(r)$ behave like the current and voltage on a uniform line.

The equivalent bi-conical line of semi-angle A has a characteristic impedance $K=120 \log_e \cot \frac{1}{2}A$ ohms.

Schelkunoff's original paper concerned cones with semi-angles less than 1.8°. In a later paper he has obtained solutions of Maxwell's equations for wide-angle cones.[8] The formulation is slightly different there, but sets of equations are developed which are essentially the same as Eqs. (3), (4), and (5) in this paper. He points out methods of obtaining approximate solutions which apply when the cone angle is very small, and when it approaches 90°, and also mentions that it would be feasible to carry out the approximations to the second order in $1/K$ if desired. However, because of the practical importance of wide-angle cones, with semi-angles from 20° to 60°, it was felt to be desirable to break away completely from any approximations that were tied to the ends of the range of semi-angle. This was done by showing that it is possible to evaluate the ratio $I_{nr} \cdot I_{nk}/I_{nn}'$ that occurs in the theory without evaluating the special solutions of Legendre's equation, $L_n(\cos\theta)$, that occur in the integrals I_{nr}, I_{nk}, and I_{nn}'.

[8] Schelkunoff, Proc. I.R.E. **34**, 29 (1946).

SOLUTIONS OF MAXWELL'S EQUATIONS

The principal waves for the space between the cones are given by the formulae:

$$r \cdot E_\theta = z_0/\sin\theta \cdot \{(1+Ky_t)e^{-j\beta(r-l)} + (1-Ky_t)e^{j\beta(r-l)}\},$$

$$r \cdot H_\phi = 1/\sin\theta \cdot \{(1+Ky_t)e^{-j\beta(r-l)} - (1-Ky_t)e^{j\beta(r-l)}\},$$

$$z_0=(\mu/\epsilon)^{\frac{1}{2}} \quad \beta^2=\omega^2\mu\epsilon \quad K=120 \log_e \cot \tfrac{1}{2}A \text{ ohms.}$$

In order not to introduce a new set of symbols, the form of Maxwell's equations given by Schelkunoff[3] as formulae (80) will be used. These are:

$$r \cdot E_\theta = -\partial V/\partial\theta, \quad j\omega\epsilon \cdot r^2 \cdot E_r = -n(n+1)A',$$

$$r \cdot H_\phi = -\partial A'/\partial\theta, \quad V=-1/j\omega\epsilon \cdot \partial A'/\partial r. \tag{1}$$

When the dimensions of the feeding transmission line are small enough, the function A' for the space between the cones becomes:

$$A'=\sum_n a_n/2\pi \cdot 1/n(n+1) \cdot \frac{(\beta r)^{\frac{1}{2}} \cdot J_{n+\frac{1}{2}}(\beta r)}{(\beta l)^{\frac{1}{2}} \cdot J_{n+\frac{1}{2}}(\beta l)} \cdot L_n(\theta) \tag{2a}$$

$L_n(\theta)$ is a solution of Legendre's equation that vanishes at $\theta=A$ and at $\theta=\frac{1}{2}\pi$. The number n is real, but not an integer in all cases.

In the space outside the boundary sphere the solutions must represent outward-moving waves. In this region

$$A'=\sum_{k=1,3,5,\ldots} b_k/2\pi \cdot 1/k(k+1) \cdot \frac{(\beta r)^{\frac{1}{2}} \cdot H_{k+\frac{1}{2}}^{(2)}(\beta r)}{(\beta l)^{\frac{1}{2}} \cdot H_{k+\frac{1}{2}}^{(2)}(\beta l)} \cdot P_k(\theta). \tag{2b}$$

$P_k(\theta)$ is the Legendre's polynomial, and the values of k are integers because this field ex-

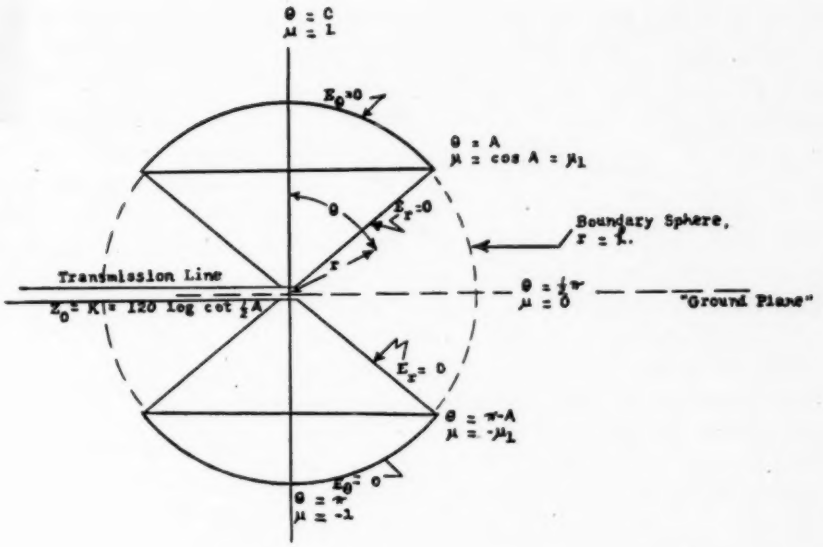


FIG. 2. The geometry of the wide-angle dipole—the feeding transmission line has negligible transverse dimensions. This diagram defines r, θ, μ, A, and $\mu_1 = \cos A$, and also the "boundary sphere."

tends from $\theta=0$ to $\theta=\pi$. Only odd integers are needed in the problem of the symmetrical dipole, because the radial field vanishes on the plane $\theta=\frac{1}{2}\pi$.

The fields described by Eqs. (1), (2a), and (2b) are solutions of Maxwell's equations and fit certain boundary conditions. The radial electric field vanishes on the sides of the cones and on the ground plane, the radial E field is also finite at the origin, and the exterior field is composed of outward-traveling waves. It remains to make the tangential fields vanish over the end surfaces of the dipole, and to make the inside and outside fields fit on the boundary sphere between $\theta=A$ and $\theta=\pi-A$. This result is obtained by using the orthogonal properties of the functions $dP_k/d\theta$ and of the functions $L_n(\theta)$. In this way the fields E_r and H_ϕ for the inside space are made to fit the outside fields on the boundary sphere between $\theta=A$ and $\theta=\pi-A$, and then the outside E_θ field is made to fit the inside field between the cones and vanish over the end surfaces of the cones.

There is a difficulty in that the field at the corner where the conical and spherical surfaces meet must be infinite, and can hardly be represented by a finite number of terms. Actually, the corner cannot be perfectly sharp, and it will probably be practical to represent the field by a finite number of terms.

Since the H_ϕ fields must match at $r=l$, it follows that*

$$r\cdot H_\phi = -\tfrac{1}{2}\pi\Sigma_k b_k/k(k+1)\cdot\partial P_k/\partial\theta$$
$$= 2/\sin\theta\cdot K\cdot y_t - \tfrac{1}{2}\pi\Sigma_n a_n/n(n+1)\cdot\partial L_n/\partial\theta.$$

Integrating this equation between the limits A and $\pi-A$ gives

$$4\log_e\cot\tfrac{1}{2}A\cdot K\cdot y_t = 2/2\pi\Sigma_k b_k/k(k+1)P_k(\cos A)$$

or

$$4\pi y_t = 120/K^2\cdot\Sigma_k b_k/k(k+1)P_k(\cos A). \quad (3)$$

This shows that each term in the external field can be considered as contributing to the "terminating admittance." It does not mean, of course, that the external field is independent of the internal field, because the coefficients b_k can be transformed into the coefficients a_n by a linear set of equations.

Since the E_r fields must match at the boundary sphere when $A<\theta<\pi-A$,

$$l\cdot E_r = \tfrac{1}{2}\pi\Sigma_n a_n\cdot L_n(\theta) = \tfrac{1}{2}\pi\Sigma_k b_k\cdot P_k(\theta).$$

The functions $L_n(\theta)$ are orthogonal in the range $\cos A>\mu>\cos(\pi-A)$, so multiplying by $L_m(\mu)\cdot d\mu$ and integrating over this range gives

$$a_m\cdot\int_{-\mu_1}^{\mu_1} L_m^2\cdot d\mu = \Sigma_k b_k\cdot\int_{-\mu_1}^{\mu_1} P_k\cdot L_m\cdot d\mu. \quad (4)$$

In the limiting case of the thin dipole, the num-

* Throughout this paper, $\mu=\cos\theta$, $\mu_1=\cos A$, $P_k(\theta)=P_k(\mu)$, and $L_n(\theta)=L_n(\mu)$.

bers n approach the values 1, 3, 5, $\cdots$ etc., and the functions $L_n(\mu)$ become very like the functions $P_1(\mu)$, $P_3(\mu)$, $P_5(\mu)$ $\cdots$ etc., except at the ends of the range. Under these conditions

$$\operatorname*{Lim}_{A\to 0} a_m = a_k = b_k. \tag{4a}$$

The same set of Eqs. (4) can also be obtained by multiplying the values of the H_ϕ fields on the boundary surface by $\sin\theta \cdot dL_m/d\theta \cdot d\theta$ and integrating from A to $\pi - A$. Thus if a sufficient number of coefficients a_m are used, these equations imply that both E_r and H_ϕ fields will be matched at the boundary surface.

The values of the E_θ fields at $r = l$ are obtained from the basic equations (1), (2), and (2a), and the equations for the principal waves. Thus, near the sphere through the ends,

$$r \cdot E_\theta \text{ inside} = \frac{2z_0}{\sin\theta} + \frac{1}{2\pi j\omega\epsilon} \cdot \Sigma_n a_n/n(n+1) \cdot \frac{\dfrac{d}{dl}[(\beta l)^{\frac{1}{2}} \cdot J_{n+\frac{1}{2}}(\beta l)]}{(\beta l)^{\frac{1}{2}} J_{n+\frac{1}{2}}(\beta l)} \cdot dL_n/d\theta,$$

$$r \cdot E_\theta \text{ outside} = \frac{1}{2\pi j\omega\epsilon} \cdot \Sigma_k b_k/k(k+1) \cdot \frac{\dfrac{d}{dl}[(\beta l)^{\frac{1}{2}} \cdot H_{k+\frac{1}{2}}^{(2)}]}{(\beta l)^{\frac{1}{2}} H_{k+\frac{1}{2}}^{(2)}} \cdot dP_k/d\theta.$$

To reduce the amount of writing, the following notation will be used:

$$S_n = (\beta l)^{\frac{1}{2}} J_{n+\frac{1}{2}}(\beta l), \qquad S_n' = d/d(\beta l) S_n, \quad z_0^2 = \mu/\epsilon,$$
$$R_k = (\beta l)^{\frac{1}{2}} H_{k+\frac{1}{2}}^{(2)}(\beta l), \quad R_k' = d/d(\beta l) R_k, \quad \beta^2 = \omega^2\mu\epsilon.$$

Since the inside and outside fields must match at the boundary sphere, between the angles A and $\pi - A$, in this range

$$4\pi j z_0/\sin\theta + z_0 \Sigma_n a_n/n(n+1) \cdot S_n'/S_n \cdot dL_n/d\theta = z_0 \Sigma_k b_k/k(k+1) \cdot R_k'/R_k \cdot dP_k/d\theta. \tag{5}$$

Over the ends of the dipoles, the E_θ field is tangential to the metallic surface, and therefore vanishes. The functions $dP_k/d\theta$ have certain orthogonal properties that allow the coefficients b_k to be fixed in such a way that the tangential field vanishes over the ends of the cones, and the fields match over the boundary sphere between the cones. The coefficients b_k here play the same sort of part as the coefficients in a Fourier's series, and the external field is being made to fit an arbitrary function at the boundary sphere. The arbitrary function is one which is zero over the ranges of $\theta = 0$ to A, and $\pi - A$ to π, and is equal to the inside field between the cones.

From the section on "integrals used involving Legendre's functions," it follows that

$$\int_0^\pi \sin\theta \cdot dP_k/d\theta \cdot dP_r/d\theta \cdot d\theta = 0 \text{ if } k \neq r, \text{ and } = 2\frac{r^2+r}{2r+1} \text{ if } k = r. \tag{6a}$$

Also

$$\int_A^{\pi-A} \sin\theta \cdot dL_m/d\theta \cdot dP_k/d\theta \cdot d\theta = k(k+1) \int_{-1}^{\mu_1} L_m \cdot P_k \cdot d\mu. \tag{6b}$$

Multiplying Eq. (5) by $\sin\theta \cdot dP_r/d\theta \cdot d\theta$ and integrating from 0 to π gives

$$z_0 \cdot b_r/r(r+1) \cdot R_r'/R_r \frac{2r(r+1)}{2r+1} = 2\pi j \int_A^{\pi-A} (r \cdot E_\theta \text{ inside}) \sin\theta \cdot dP_r/d\theta \cdot d\theta.$$

The above equation can be rewritten

$$P_r(\cos A) = \tfrac{1}{4}\pi \cdot 1/(2r+1) \cdot jR_r'/R_r \cdot b_r - \tfrac{1}{4}\pi \cdot (r^2+r) \Sigma_n a_n/n(n+1) \cdot jS_n'/S_n \cdot \int_0^{\mu_1} L_n \cdot P_r \cdot d\mu. \tag{6c}$$

This set of equations, therefore, fixes the coefficients b_k in such a way that the field vanishes over the ends of the dipoles, and the outside field matches the inside E_θ field over the boundary surface. Introducing the values of a_n from the set of Eqs. (4) introduces the conditions that the E_r and H_ϕ fields shall also match on the boundary surface. Therefore, if the set of numbers b_k is determined from the set of Eqs. (7)

that follow, the boundary conditions will be satisfied if an infinite number of terms are used.

From Eqs. (4) and (6c) it follows that

$$P_r(\cos A) = \tfrac{1}{4}\pi \cdot 1/(2r+1) \cdot jR_r'/R_r \cdot b_r - \Sigma_k \Sigma_n \tfrac{1}{4}\pi \cdot (r^2+r)/(n^2+n) \cdot jS_n'/S_n \cdot I_{nr} \cdot I_{nk}/I_{nn}' \cdot b_k, \quad (7)$$

where

$$I_{nr} = \int_0^{\mu_1} L_n \cdot P_r \cdot d\mu = \frac{(1-\mu_1^2) P_r(\mu_1) \cdot dL_n/d\mu_1}{(r^2+r-n^2-n)},$$

(see reference 9)

$$I_{nk} = \int_0^{\mu_1} L_n \cdot P_k \cdot d\mu,$$

$$I_{nn}' = \text{Lim as } e \text{ tends to } 0 \int_0^{\mu_1} L_n \cdot L_{n+e} \cdot d\mu$$

$$= \text{Lim as } e \text{ tends to } 0 \frac{(1-\mu_1^2) L_{n+e} dL_n/d\mu_1}{(2n+1)e+e^2}$$

$$= \frac{(1-\mu_1^2) dL_n/dn \cdot dL_n/d\mu_1}{2n+1}.$$

The fact that the tangential field vanishes accounts for the fact that the limits of integration appear as $\mu_1 = \cos A$, and the half-range has been used because of the antisymmetry of the functions.

The integral I_{nn}' can be further transformed by virtue of the fact that $L_n(\mu_1) = 0$ by definition. If L_n is considered as a function of μ_1 and n, then for all $\mu_1 dL_n = 0$. That is to say,

$$dL_n(\mu_1) = \partial L_n/\partial \mu_1 \cdot d\mu_1 + \partial L_n/\partial n \cdot dn = 0, \text{ or}$$

$$dL_n/dn = -(\partial L_n/\partial \mu_1)/(\partial n/\partial \mu_1).$$

Here n is regarded as a function of μ_1, and from this it follows that

$$I_{nn}' = \frac{(1-\mu_1^2) dL_n/d\mu_1 \cdot dL_n/dn}{(2n+1)}$$

$$= -\frac{(1-\mu_1^2)}{(2n+1)} (dL_n/d\mu_1)^2 \cdot d\mu_1/dn. \quad (8)$$

[9] E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Teddington, England, 1931), p. 38, Ex. 9.

Thus the ratio appearing in formula (7) becomes

$$I_{nr} \cdot I_{nk}/I_{nn}' = -\frac{(2n+1) \cdot (1-\mu_1^2) \cdot dn/d\mu_1 \cdot P_r(\mu_1) \cdot P_k(\mu_1)}{(r^2+r-n^2-n)(k^2+k-n^2-n)}. \quad (9)$$

Making use of this result in Eq. (7), and also writing for brevity

$$W_r = 1/(2r+1) \cdot jR_r'/R_r,$$

$$s_n = (2n+1)/(n^2+n) \cdot (1-\mu_1^2) \cdot dn/d\mu_1 \cdot jS_n'/S_n,$$

$$(r, n) = (r^2+r-n^2-n)$$

gives the following set of linear equations for b_k:

$$4\pi \cdot P_r(\mu_1) = b_r \cdot W_r + \Sigma_n \Sigma_k b_k \cdot \frac{(r^2+r) P_r(\mu_1) \cdot P_k(\mu_1) \cdot s_n}{(r, n)(k, n)}. \quad (10)$$

In addition to this set, it has been shown that

$$4\pi \cdot K^2/120 \cdot y_t = \Sigma_r b_r \cdot P_r(\mu_1)/(r^2+r). \quad (3)$$

It is evident that to determine an approximate value for the effective terminating admittance, y_t, it is necessary to select a number of the Eqs. (10) equal to the selected number of coefficients b_k, and to eliminate the coefficients b_k between this set and Eq. (3). The numbers n are the roots of the equation $L_n(A) = 0$.

If it is assumed that it is sufficient to take the first, or smallest n value, and the first two k values, namely, 1 and 3, then the set becomes:

$$4\pi \cdot P_1 = b_1 \{W_1 - 2P_1^2 \cdot s_n/(1, n)^2\} + b_3 \left\{\frac{2P_1 \cdot P_3 \cdot s_n}{(1, n)(3, n)}\right\},$$

$$4\pi \cdot P_3 = b_1 \left\{\frac{12 P_3 \cdot P_1 \cdot s_n}{(3, n)(1, n)}\right\} + b_3 \left\{\frac{W_3 + 12 P_3^2 s_n}{(3, n)^2}\right\},$$

$$4\pi K^2/120 \cdot y_t = b_1 \cdot P_1/2 + b_3 \cdot P_3/12,$$

where

$$P_1 = P_1(\cos A) \quad \text{and} \quad P_3 = P_3(\cos A).$$

By eliminating the coefficients b_k, and assuming that a sufficiently good approximation can be obtained by using the T.E.M. and one other inside wave, and the first two outside waves,

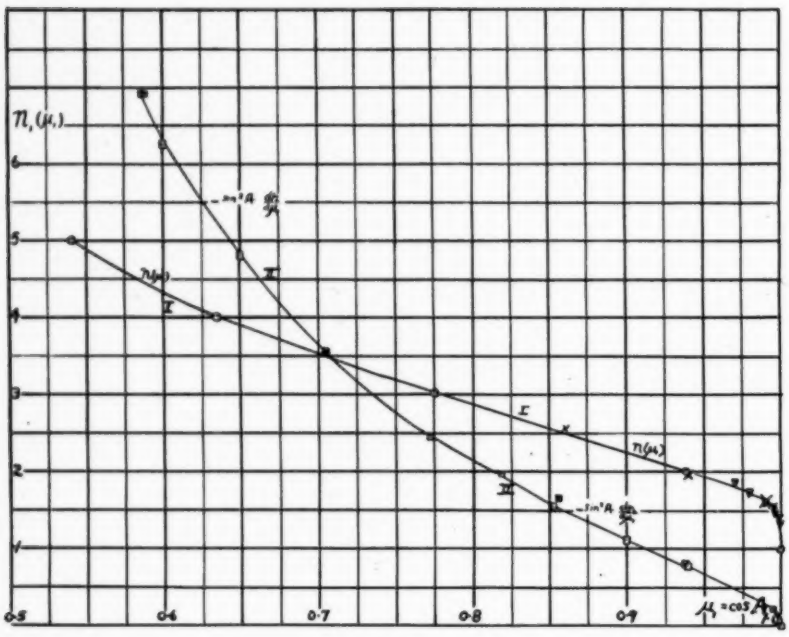


FIG. 3. The functions n_1 and $-\sin^2 A \cdot dn/d\mu_1$—curve I is $n_1(\mu_1)$. This is the first root in n of $L_n(\mu_1)=0$. (L_n satisfies Legendre's equation and vanishes at $\mu=0$.) The points marked as circles are the roots of $P(\mu_1)_{2k+1}$ and $Q(\mu_1)_{2k}$. The crosses are from Eq. (13a), the triangles from $n=1+1/(\log_e \operatorname{cosec} A-1)$. Curve II is $-\sin^2 A\, dn/d\mu_1$. The open squares are from Eq. (15), the shaded squares from the mean slope of the segments joining the circles of curve I.

the following formula for y_t is obtained:

$$K^2 \cdot y_t/120 = \frac{\frac{1}{2}P_1^2/W_1 + 1/12 \cdot P_3^2/W_3 + \frac{1}{6}[1/(1,n) - 6/(3,n)]^2 P_1^2 \cdot P_3^2 s_n/(W_1 W_3)}{1 + 2P_1^2/(1,n)^2 \cdot 1/W_1 \cdot s_n + 12 P_3^2/(3,n)^2 \cdot 1/W_3 \cdot s_n},$$

where

$$1/W_r = (2r+1)R_r/jR_r',$$

$$s_n = (2n+1)/(n^2+n) \cdot (1-\mu_1^2) dn/d\mu_1 \cdot (jS_n'/S_n),$$

$$(r,n) = (r^2+r-n^2-n)$$

$$P_1^2 = (P_1(\cos A))^2 \quad P_3^2 = (P_3(\cos A))^2. \tag{11}$$

It is interesting to note the limiting form taken by formula (11) as A tends to 0. In this case μ_1 tends to 1, n tends to 1, $1/(1,n) \to -\frac{1}{3}(n-1)$, and $\sin^2 A\, dn/d\mu_1 \to -(n-1)^2$, so that**

$$\lim_{A \to 0} s_n = -\tfrac{3}{2} \cdot (n-1)^2 \cdot (jS_1'/S_1).$$

$1/(3,n)$ can be neglected compared to $1/(1,n)$ in the numerator. Thus,

$$\lim_{A \to 0} K^2 \cdot y_t/120 = \frac{\frac{1}{2} \cdot 1/W_1 + 1/36 \cdot 1/W_3 \cdot (3 - (jS_1'/S_1)/W_1)}{1 - \frac{1}{3}(jS_1'/S_1)/W_1}$$

$$= \tfrac{3}{2} \cdot 1/(jR_1'/R_1 - jS_1'/S_1) + 1/12 \cdot 7R_3/jR_3'$$

$$= \tfrac{3}{2} \cdot 1/(jR_1'/R_1 - jS_1'/S_1) + \frac{7/12 \cdot (1 - R_3S_3'/R_3'S_3)}{jR_3'/R_3 - jS_3'/S_3}$$

$$= \tfrac{1}{2}\pi L(\tfrac{3}{2} J_{1\frac{1}{2}} \cdot H_{1\frac{1}{2}}^{(2)} + 7/12 \cdot J_{3\frac{1}{2}} \cdot H_{3\frac{1}{2}}^{(2)} \cdot (1 - R_3S_3'/R_3'S_3)).$$

And so to this approximation the "terminating admittance" of a thin dipole is given by

$$K^2 \cdot y_t = 60\pi L(\tfrac{3}{2} J_{1\frac{1}{2}} \cdot H_{1\frac{1}{2}}^{(2)} + 7/12 \cdot J_{3\frac{1}{2}} \cdot H_{3\frac{1}{2}}^{(2)} \cdot (1 - R_3S_3'/R_3'S_3)), \tag{12}$$

whereas the Schelkunoff formula would give (formula 18[3])

$$K^2 y_t = 60\pi L(\tfrac{3}{2} J_{1\frac{1}{2}} \cdot H_{1\frac{1}{2}}^{(2)} + 7/12 \cdot J_{3\frac{1}{2}} \cdot H_{3\frac{1}{2}}^{(2)} + \cdots).$$

** Footnote: It can be shown that $(jR_k'/R_k - jS_k'/S_k) = \pi L/2 \cdot J_{k+\frac{1}{2}} \cdot H_{k+\frac{1}{2}}^{(2)}$, by using the relations $H_{p-1}^{(2)} \cdot J_p - H_p^{(2)} \cdot J_{p-1} = 2/j\pi L$ and $Z_p' = -p/LZ_p + Z_{p-1}$.

Thus the Eqs. (10) for b_k and the approximate formula (11) for y_t do not depend on definite integrals that involve the functions $L_n(\theta)$, but instead contain the functions s_n that depend upon the way in which the numbers n vary with $\cos A$. Therefore, the first step in calculating the admittance y_t is to determine the values of n and $dn/d\mu_1$ as functions of $\cos A$, and for this purpose a brief examination of the properties of the $L_n(\theta)$ functions is introduced.

SOME PROPERTIES OF THE $L_n(\theta)$ FUNCTIONS

The $L_n(\theta)$ functions were defined as those solutions of Legendre's differential equation

$$d/d\mu(1-\mu^2)dL_n/d\mu+n(n+1)L_n=0, \quad \mu=\cos\theta,$$

that vanish both at $\theta=\frac{1}{2}\pi$ and some angle A. For any given angle A there will be a set of numbers $n_1, n_2, \cdots n_j, \cdots$ that satisfy $L_n(A)=0$. The E_r field corresponding to one of these values n_j has j maxima of intensity between A and $\frac{1}{2}\pi$. Conversely, if j is fixed, n_j will be a continuous function of A, and will be integral for certain values of A. By examination of the curves for P_{2k+1} and Q_{2k} it is apparent that these functions meet the conditions for $L_n(\theta)$, functions for certain particular values of A. In these cases n_j is an even or odd integer, and j is the number of maxima and minima between $\frac{1}{2}\pi$ and A.

In this problem the only j value considered is 1. The integral values of $n_1(\cos A)$ are plotted as functions of $\mu_1=\cos A$ as small circles in curve I of Fig. 3. For values of μ_1 that are close to 1, it can be shown that $n=1+1/(\log_e \operatorname{cosec}\frac{1}{2}A-1)$, a result easily found from Schelkunoff's equations (114).[3] These are the points plotted as triangles. The points plotted as crosses are the roots in n of the equation

$$0=\sum_{j=1,3}(2j+1)/(j,n)\cdot P_j^2(\mu_1)$$

which, as will be apparent, should give the same curve. Thus there are three methods of obtaining points on the curve for the first root n_1 of $L_n(\mu_1)=0$ as a function of μ_1. If it was desired to obtain a result involving the second root, n_2, as well, this could be obtained from the roots of $P_1, P_3 \cdots$ and $Q_2, Q_4 \cdots$ as well as from the formula above. This was not necessary in this problem.

While the L_n functions are infinite at the points $\mu=1$ and $\mu=-1$, the actual fields vanish in the metal of the conductors. In fact, this suggests that if an expansion of the L_n functions in terms of the P_k functions were attempted, it would only be necessary to make the expansion equal to the L_n functions in the range of μ that lies between the cones, and let it be equal to zero in the ranges corresponding to the metal of the conductors. This procedure avoids certain difficulties that arise if the attempt is made to make the expansion equal to L_n over the whole range. The expansion is carried out in the following manner.

Suppose

$$\Sigma_j a_j\cdot P_j(\mu)=L_n(\mu)$$

between μ_1 and $-\mu_1$. Then

$$a_j=\tfrac{1}{2}(2j+1)\int_{-\mu_1}^{\mu_1}L_n(\mu)\cdot P_j(\mu)\cdot d\mu.$$

The integral is taken over the range $-\mu_1$ to μ_1 only, because it has been stated that the sum vanishes outside this range. Hence it follows that

$$a_j=\tfrac{1}{2}(2j+1)\cdot\frac{2\sin^2 A\cdot P_j(\mu_1)\cdot dL_n/d\mu_1}{(j^2+j-n^2-n)},$$

and between μ_1 and $-\mu_1$

$$L_n(\mu)/(\sin^2 A\cdot dL_n/d\mu_1)=\sum_{j\ \text{odd}}(2j+1)/(j,n)\cdot P_j(\mu_1)\cdot P_j(\mu). \quad (13)$$

If Legendre's equation is differentiated with regard to n, then it appears that

$$d/d\mu[(1-\mu^2)d/d\mu(\partial L_n/\partial n)]+n(n+1)\partial L_n/\partial n=-(2n+1)L_n.$$

If $\Sigma b_j\cdot P_j(\mu)$ is a function that vanishes between μ_1 and 1, and between $-\mu_1$ and -1, but between μ_1 and $-\mu_1$ equals $\partial L_n/\partial n$, then in the range μ_1 to $-\mu_1$

$$\Sigma_j(n^2+n-j^2-j)\cdot b_j\cdot P_j=-(2n+1)\Sigma_j a_j\cdot P_j,$$

so that

$$b_j=(2n+1)/(j^2+j-n^2-n)a_j,$$

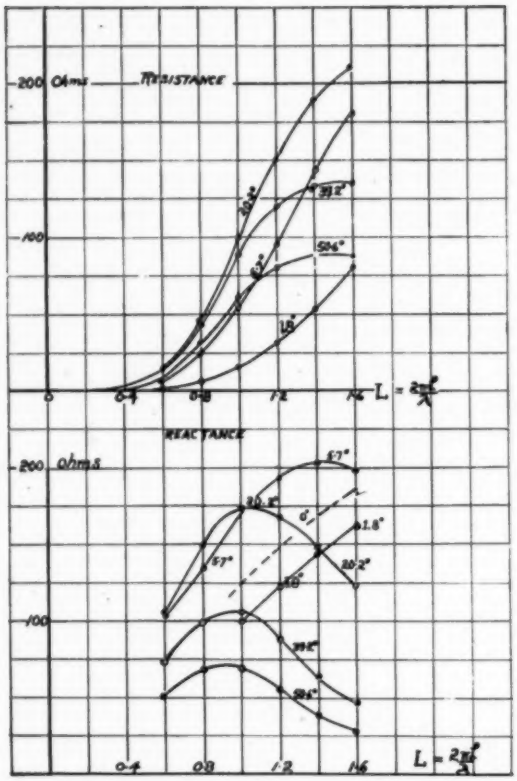


FIG. 4. Formula (11)—approximate values of the impedance $K^2 \cdot y_i$ at a point where the line length plus dipole radius together make $\frac{1}{4}\lambda$, the line's characteristic impedance being $K = 120\log_e \cot\frac{1}{2}A$. The dotted curves represent Schelkunoff's results for thin cones.

and, using the value of a_j given above,

$$\partial L_n/\partial n = \sin^2 A\, dL_n/d\mu_1 \cdot \sum_{j\ \text{odd}} (2n+1)\times(2j+1)/(j^2+j-n^2-n)^2 \cdot P_j(\mu_1)P_j(\mu) \quad (14)$$

in the range μ_1 to $-\mu_1$.

Since n and μ_1 are so related that $L_n(\mu_1)=0$,

$$dn/d\mu_1 = -\partial L_n/\partial\mu_1/\partial L_n/\partial n$$

and the important quantity $\sin^2 A \cdot dn/d\mu_1$ is given by:

$$(2n+1)\cdot\sin^2 A \cdot dn/d\mu_1 = -1/(\Sigma_j(2j+1)/(j,n)^2 \cdot P_j^2(\mu_1)). \quad (15)$$

Returning to Eq. (13), we note that the sum

$$S(n,\mu_1,\mu) \equiv \sum_{j\ \text{odd}} (2j+1)/(j,n)\cdot P_j(\mu_1)\cdot P_j(\mu)$$

is only equal to

$$L_n(\mu)/(\sin^2 A : dn/d\mu_1)$$

when n is one of the roots of $L_n=0$. The points plotted as crosses in Fig. 3 were found by trying different values of n in the equation

$$0 = S_7(n,\mu_1) \equiv (3P_1^2/(1,n)+7P_3^2/(3,n) +11P_5^2/(5,n)+15P_7^2/(7,n)), \quad (13a)$$

where $P_1^2=(P_1(\mu))^2$, etc. This formula does not give results of the highest accuracy, but it serves as a check on the validity of interpolating between the values obtained by plotting the roots in n of $P_n(\mu_1)$ and $Q_n(\mu_1)$.

In the curves for $-\sin^2 A \cdot dn/d\mu_1$ (curve II, Fig. 3) the points marked as open squares were obtained from formula (15). The shaded squares are derived from the slope of the segments joining the P_n and Q_n points of curve I.

From the Eqs. (114) in Schelkunoff's paper,[3] it is easy to show that

$$\operatorname{Lim}_{A\to 0}(-\sin^2 A \cdot dn/d\mu_1) = (n-1)^2 = D^2.$$

This result can be used at small cone angles. It is used in deriving formula (12).

INTEGRALS USED INVOLVING LEGENDRE'S FUNCTIONS

If T_n and T_m are any two solutions of Legendre's equation, where n and m have any real values, it can be shown by integrating by parts that

$$\int_A^{\pi-A} \sin\theta \cdot dT_n/d\theta \cdot dT_m/d\theta \cdot d\theta = [\sin\theta \cdot T_m(\theta)\cdot dT_n(\theta)/d\theta]_A^{\pi-A} + n(n+1)\int_{-\mu_1}^{\mu_1} T_n \cdot T_m \cdot d\mu \quad (16)$$

and also, either by integrating by parts or by reference to Hobson,[9] it is easy to see that

$$\int_{-\mu_1}^{\mu_1} T_n \cdot T_m \cdot d\mu = \frac{\sin^2 A}{(n^2+n-m^2-m)} \cdot [T_n \cdot dT_m/d\mu - T_m \cdot dT_n/d\mu]_{-\mu_1}^{\mu_1}. \quad (17)$$

These two formulae are useful in evaluating the integrals that appear in this report. (As before, μ_1 equals $\cos A$.)

From the first of these formulae it is apparent that the functions $dL_n/d\theta$ are orthogonal in the

range A to $\pi-A$, and the functions $dP_k/d\theta$ are orthogonal in the range 0 to π. The second shows that the functions L_n are orthogonal in the range A to $\pi-A$, and the P_k functions in the range, 0 to π, or -1 to 1 in μ. This orthogonality permits the fields to be fitted to arbitrary functions at the boundary $r=l$.

THE DEGREE OF APPROXIMATION OBTAINED

It has been shown that the effective admittance terminating a bi-conical dipole can be obtained by eliminating the set of coefficients b_k from the equations

$$4\pi \cdot P_r(\mu_1) = W_r \cdot b_r - \Sigma_n \Sigma_k (r^2+r) \cdot \frac{P_r(\mu_1) \cdot P_k(\mu_1)}{(r,n)(k,n)} \cdot s_n \cdot b_k, \quad (10)$$

$$4\pi K^2/120 \cdot y_t = \Sigma_r P_r(\mu_1)/(r^2+r) \cdot b_r. \quad (3)$$

In deriving Eq. (11) from this set of equations the assumption that only $n=n_1(\mu_1)$ and $k=1$ and $k=3$ need be considered was introduced. To justify this assumption it would be desirable to carry the approximation to the next stage, and see whether the effect of using two n values, $n_1(\mu_1)$ and $n_2(\mu_1)$, was important, or whether the effect on the result of another k value, that is, k taking the values 1, 3, and 5, was important.

This has not been done because of the complexity of the formula in the stage involving only $n_1(\mu_1)$, 1, and 3. The result therefore depends upon experimental verification to check its usefulness.

However, it is not difficult to reduce the degree of approximation by neglecting all the inside waves except the T.E.M. waves, and then it is possible to see how big a change occurs in the value of $K^2 \cdot y_t$ when this $n_1(\mu_1)$ term is removed. When this is done, it is found that there is a very large difference at small cone angles, but that this difference decreases as the cone angle increases.

Combining Eq. (3) with Eq. (7), and neglecting all terms involving n, gives the simple formula

$$K^2 \cdot y_t = 120\Sigma_r (2r+1)/(r^2+r) \cdot P_r^2(\mu_1) \cdot R_r/jR_r'. \quad (18)$$

This is a formula which takes complete account of the fact that the tangential field must vanish over the ends of the dipole, but does not match the E_r fields on the spherical boundary surface between the ends of the dipole.

The curves found by formula (11) are plotted in Fig. 4. The curves based on this formula (18) are plotted as solid lines in Fig. 5, while certain curves based on formula (11) are plotted as dotted lines for comparison. The sum in formula (18) was carried over values of r of 1, 3, 5, and 7.

It is clear that for the thin dipole, (0°), formula (18) gives completely different results from formula (11). At 20.2° semi-angle the two formulae are of the same order of magnitude, and at 50.6° they agree to better than 25 percent over the whole range of $L=2\pi l/\lambda$ from 0.6 to 1.6.

Formula (18) has been carried up to $r=7$, although formula (11) only includes $k=1$ and 3. These extra terms have a negligible effect on the resistance in formula (18), but at small angles

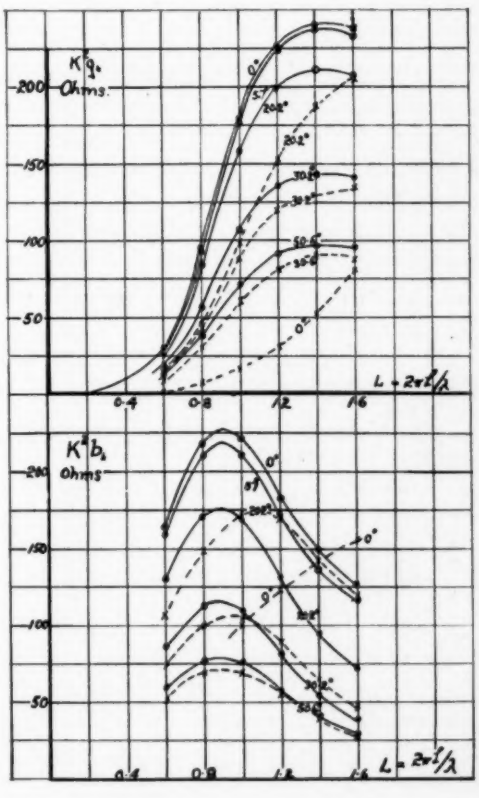


FIG. 5. Formula (18)—results obtained if the higher order inside waves are neglected, measured ¼λ from the ends of the dipole, according to formula (18). The results of formula (11) shown as dotted lines for comparison.

they amount to 20 percent of the reactance by this formula. Their effect is much smaller at wide cone angles, and is less than 6 percent of the reactance at $L=1.6$ for the dipoles that are wider than 20°.

In general, the resistance value is probably quite good from formula (11), but the reactance value is perhaps only to be expected to be within 10 percent, for the thin cones.

If rough results are all that is required, the simple formula (18) will probably be within 30 percent for this range of L. Calculation by formula (18) is far less laborious than by formula (11), but can only be applied to cones whose semi-angles exceed 39.2°.

Considering the limiting case where A tends to 90°, it can be seen that the individual terms of formula (18) are identical with the terms Y_n given by Stratton and Chu[4] for the terminal admittance of a spherical dipole, because

$$(P_n^1(0))^2 = \text{limit as } A \text{ tends to } 90° \left[\frac{P_n(A)}{\log_e \cot \tfrac{1}{2}A}\right]^2.$$

From the sketch of the successive changes of form of the dipole, it would appear that the terminating admittance for a wide-angle conical dipole should tend to become the same as the input admittance of Stratton and Chu's sphere, because their input terminals are on the equatorial plane of the sphere. Hence the formula (18) appears to agree with the spherical case as A tends to 90°.

At the other extreme, formula (12) agrees closely with Schelkunoff's formula (18) in reference 3 as far as the terminal conductance of thin cones is concerned, but is about 15 percent low in susceptance, probably because of the neglect of the terms of higher order than 3.

APPLICATION OF THE RESULTS

The object of this analysis was to determine the value of the effective "terminating admittance" $y_t = g_t + jb_t$. The theory then indicates that to the T.E.M. waves the dipole has a characteristic admittance $K = 120 \log_e \cot \tfrac{1}{2}A$, and at points near the apex the impedance will be that of a transmission line of characteristic impedance K terminated by y_t at the end of the dipole.

If the dipole is fed by a transmission line of the same characteristic impedance K, the input admittance can be found at any position on this line by the usual method, using the transmission line formulae or impedance charts.

The theory, however, gives the impedance $K^2 \cdot y_t$ instead of y_t. This impedance is the impedance $\frac{1}{4}$-wave from the ends of the dipole, this distance including both the length of the dipole end to the junction with the transmission line, and the length along the transmission line itself.

There is always some discontinuity in line characteristics at the junction of the line and the dipole. This may not be important if the transmission line has such small transverse dimensions that they are negligible compared to a wave-length. However, at wave-lengths from 10 to 20 cm it may be difficult to achieve this condition and at the same time have the dimensions of the transmission line held closely enough to the required tolerances so that there are no irregularities in the line impedance.[10] In this case it may be necessary to measure the characteristics of the junction to see if it has negligible effect, acts as a bridged impedance, or if it has to be treated as a network with three parameters.

In this connection, the fact that it is the impedance to the principal waves that is considered is of importance, because, if the dipole is treated as a conical transmission line, and the two ends of the dipole are connected by spherical reflectors, or short circuits, at known distances l from the center, then the impedance looking outward at any point r from the center will be $jK \tan\beta(l-r)$. Now, it is consistent with this theory to consider the intersections of a sphere of radius r with the cones as the load terminals of a network, and any chosen point on the transmission line as the input terminals of the same network. Hence there exists the possibility of placing known impedances at the load terminals of the network and thereby determining its characteristics from measurements made on the input side.

These spherical reflectors or "short circuits" will not excite any modes except a reflected T.E.M. wave. The higher modes that appear when the cone is set up as a dipole penetrate into

[10] D. D. King, J. App. Phys. **17**, 844 (1946).

the space inside the boundary sphere, but, as has been shown,[3] their effect on the T.E.M. waves is the same as that of a lumped susceptance at the end of the dipole. It is only the properties of the junction with regard to T.E.M. waves that are important in this problem. Since the spherical short circuits only affect the T.E.M. waves, they will produce exactly those effects on the standing-wave pattern on the coaxial line that correspond to the transmission of T.E.M. waves through the junction. Hence characteristics of the junction net measured in this way will be the required characteristics for the transmission of T.E.M. waves. The smallest reflector must, however, have a radius great enough to avoid the distortions of field that occur near the junction.

For tests made at a single frequency with varying lengths of dipole, it would be possible to match out the discontinuities of the junction so that it is equivalent to a section of line of characteristic impedance K.

The curves of Fig. 4 give the impedance $K^2 \cdot y_t$ for a complete dipole. The case where only half a dipole is used above a ground plane is obtained by halving all the values given. If such a dipole is set up above a conducting plane, the cone should be placed so that the vertex would lie in the plane. If the transmission line is a coaxial line coming up normally through the plane, this will mean that a short length of the inner conductor will protrude until it meets the cone.

RESULTS OF THE CALCULATIONS

The results of these calculations can be presented in several forms. In Fig. 4 the resistance and reactance which should be measured at a point $\frac{1}{4}$-wave from the ends of the dipole on a line of impedance $K = 120 \log_e \cot \frac{1}{2}A$ are plotted as a function of $L = 2\pi l/\lambda$.

Figure 6 shows these values transformed to the base of the dipole. The curves for $K = 408$ ohms are taken from D. D. King's results.[10] The resistance values fit well into the family obtained from formulae (11) and (12). The fact that the resonant length is at $L = 1.46$ can be understood by reference to formulae (62) and (66) of Schelkunoff.[3] Here Schelkunoff shows that for the conical dipole the resonant length is given by

$$4l/\lambda = 1 - 97.82/K, \qquad \text{(62, reference 3)}$$

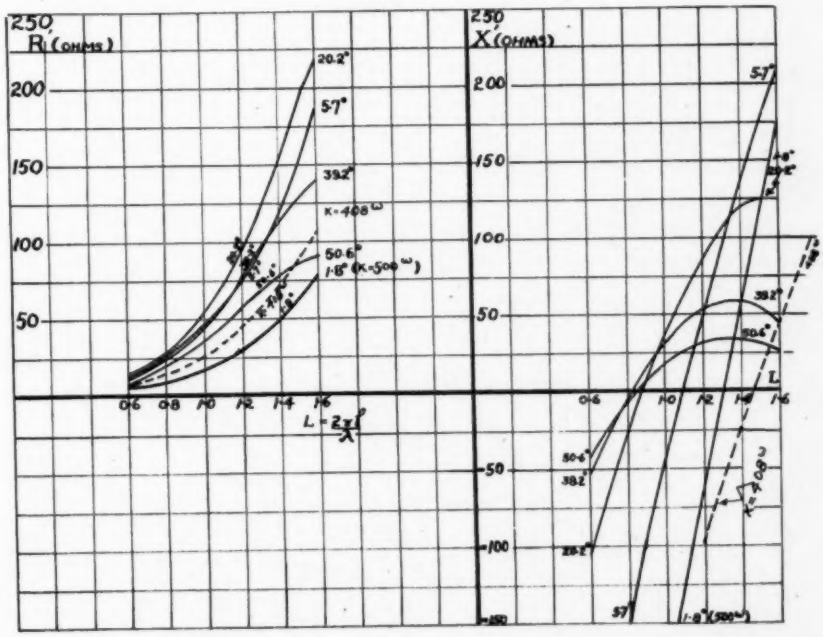


FIG. 6. The impedance at the apex of the cones—the values from formula (11) transformed by impedance chart. (The dashed curves represent D. D. King's experimental results for a cylindrical dipole of radius 6.21×10^{-3} wave-lengths, assumed to have a characteristic impedance of 408 ohms.)†

† The 1.8° curves in Figs. 6 and 7 were computed on the assumption that $n = 1$ for thin dipoles, in order to show that formula (11) agreed approximately with Schelkunoff's results for 500 ohm dipoles.
It has been noticed recently by the author that $n = 1.315$ for the 1.8° dipole, and therefore the 1.8° curves are to be regarded as a first approximation. The actual curves will be closer to the 5.7° curves.

while for the cylindrical dipole the resonant length is given by

$$4l/\lambda = 1 - 27.08/K. \qquad \text{(66, reference 3)}$$

These give the shortenings as 24 percent for the conical dipole and 6.6 percent for the cylindrical dipole when $K = 408$ ohms.

There appears to be a maximum in the input impedance of $\frac{1}{4}$-wave dipoles when the semi-angle of the cone lies between 5.7° and 20.2°. Cones of angle greater than 39.2° are coming into the region of the second resonance as the length passes $\frac{1}{4}$ wave-length. If the input impedance is plotted on a circular impedance diagram, it will be seen that the standing-wave ratio on the cones is less than 1.4 for the 39.2° and 50.6° $\frac{1}{4}$-wave dipoles. This means that there is very little reflection at the ends of the dipoles,

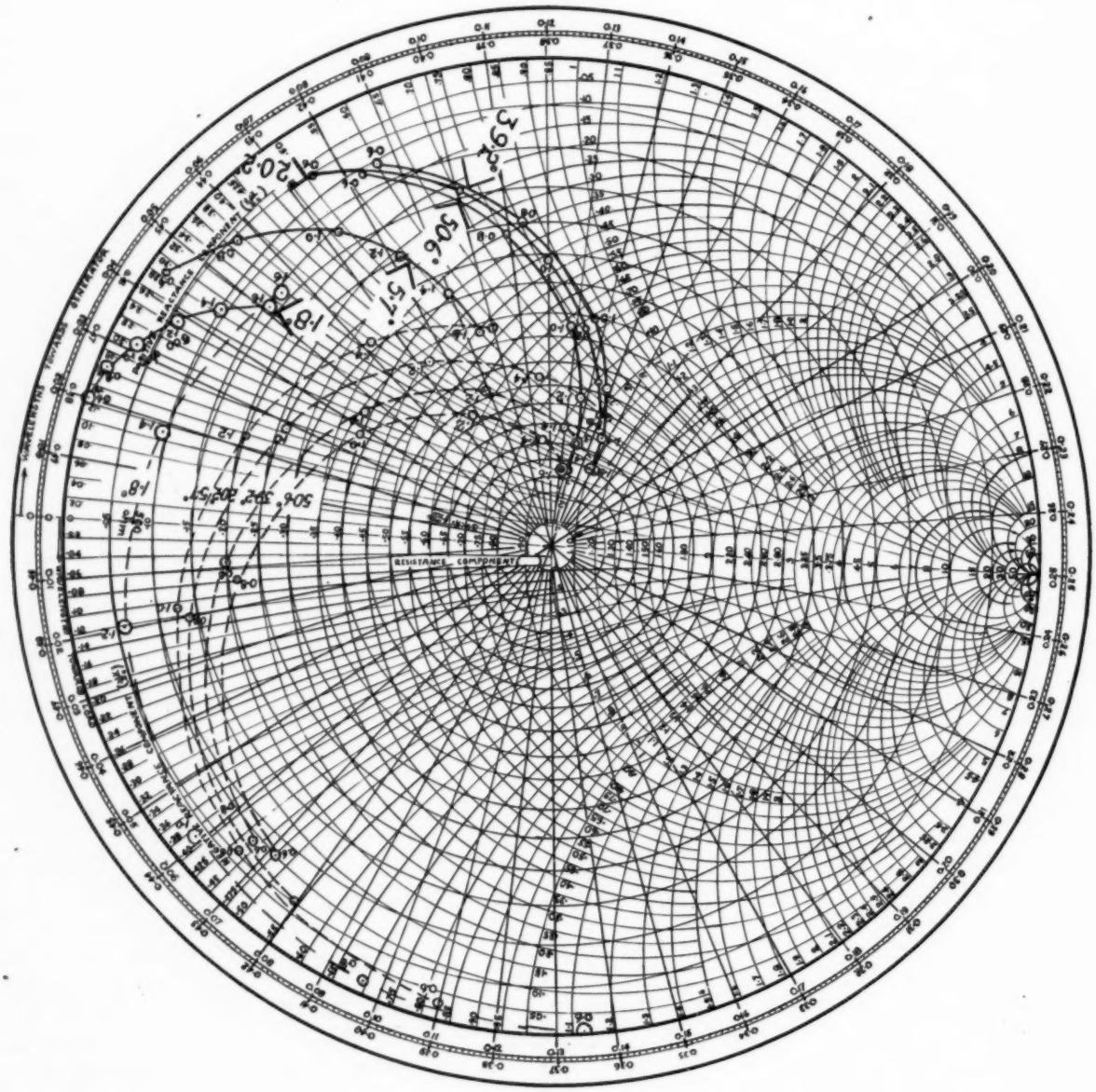


FIG. 7. Formula (11), reflection coefficient diagram—the relative end admittance, $K \cdot y_t$, (solid curves), and relative apex impedance, Z'/K, (dashed curves). The points plotted as circles represent electrical lengths of 0.6, 0.8, 1.0, 1.2, 1.4, and 1.6 radians, the semi-angles of the cones are indicated on their respective curves.†

and that the conditions are far more suitable for broad band design than they are when the dipole is narrow.

The character of these impedance curves is similar to those on page 246 of Stratton and Chu,[5] except that there is no maximum in impedance in these curves as the ratio of length to diameter of the spheroids is varied, and the resonant lengths are much shorter for cones than for the spheroids.

In some ways, the plotting of $K \cdot y_t$ on a circular impedance diagram is the most useful representation. It can be regarded either as the "admittance" at the end of the dipole, or as the impedance $\frac{1}{4}$-wave from the end of the dipole, and the "impedance" at other points, such as at the center, or some nearby position, can be obtained readily. The character of the standing-wave pattern is also evident (see Fig. 7).

CONCLUSION

A theoretical investigation has been made into the impedance of wide-angle conical dipoles. The theory covers the range of cones whose semi-angles lie between 0° and 90°. The approximate formulae obtained agree reasonably with that of Schelkunoff for the impedance of a thin conical dipole and with that of Stratton and Chu for the impedance of a spherical radiator. The curves of resistance at the center of the dipole appear to agree approximately with the measured values of D. D. King for a cylindrical dipole whose radius is 6.21×10^{-3} wave-lengths, and the difference in resonant lengths for thin dipoles between this theory and D. D. King's measurements is also reasonable in the light of certain results of Schelkunoff.

By using Maxwell's equations, and by making the tangential electric field vanish on all metallic surfaces, the general formulae (3), (4), (5), (7), and (10) were obtained. It is presumed that by eliminating the coefficients b_k from Eq. (3) and a sufficient number of Eqs. (10), solutions for y_t will be obtained that increase in accuracy as the number of terms b_k and the number of values of n are increased. Formula (11) comes by limiting k to the values 1 and 3, and taking the first value of n corresponding to the semi-angle A. This formula seems to give tolerably useful results over the whole range of A.

Formula (18) is a simpler formula which comes by neglecting the E_r fields inside the boundary sphere. The results of this formula are very different from those of formula (11) at small angles, but the difference becomes less as A approaches 90°. This formula (18) appears to be the same as that of Stratton and Chu for the spherical radiator. In this formula, the higher terms become quite small for angles of 39.2° and 50.6°.

When $K \cdot y_t$ is plotted on an impedance diagram, it becomes apparent that while the end reflections are large in thin $\frac{1}{4}$-wave dipoles, they are small in dipoles of semi-angles 30.9° and 50.6°. For these latter cases, the end reflection coefficient is less than $(0.17)^2$, corresponding to a standing-wave ratio of 1.4. This small end reflection would appear to be a physical reason why large angle cones can be matched over a wide frequency band. Because there is such a large end capacity in these cones, the resonant lengths are much reduced, and the $\frac{1}{4}$-wave dipole at 50° semi-angle is near the second resonant length.

The writer wishes to thank Professor Leopold Infeld for his very helpful discussion and criticism.

Theory of Slots in Rectangular Wave-Guides*

A. F. Stevenson
Department of Mathematics, University of Toronto, Toronto, Canada

(Received March 3, 1947)

A basic theory of slots in rectangular wave-guides is given. The analogy with a transmission line is developed and established, and detailed formulae for the reflection and transmission coefficients and for the "voltage amplitude" in the slot generated by a given incident wave are given. While the complete expressions for these quantities are quite complicated and involve the summation of infinite series, certain parts of the expressions are comparatively simple. In particular, the "resistance" or "conductance" of slots which are equivalent to series or shunt elements in a transmission line are given by fairly simple closed expressions. Guide-to-guide coupling by slots and slot arrays are also considered. A more detailed summary of the main results of the paper is given in Section 1.

1. INTRODUCTION. FUNDAMENTAL ASSUMPTIONS AND SUMMARY OF MAIN RESULTS

THE use of slots in wave-guides has proved a promising means of launching high frequency radiation. The subject has been extensively investigated by Watson.[1] In this paper, a fundamental theory of slots in wave-guides, based on the field equations, is attempted. Attention is confined, for the most part, to rectangular guides, but the principle of the method is general, and other shapes of guide could be considered if necessary.

The fundamental assumptions on which the theory is based are the following:

(1) The walls of the guide are perfectly conducting and of negligible thickness.

(2) The slot is narrow; to be more precise, we assume that

$$2\log(\text{length of slot}/\text{width of slot}) \gg 1.$$

(3) In considering the field outside the guide, the penetration of the field into the region behind the face containing the slot is neglected. In other words, we treat the problem as if the guide-face containing the slot had an infinite perfectly conducting flange on it.

(4) The guide transmits only the H_{01}-wave, and the length of the slot is near that of the first "resonance" (i.e., near $\lambda/2$).

Of these assumptions, the third is probably the one most seriously at fault.[1a] Experiment indicates that, in some cases, the penetration of the field behind the face is by no means negligible. In the case of an *array* of slots, however, the assumption is probably a good one. The fourth assumption is not in any way essential to the development of the theory, but is made because it is probably satisfied in cases of practical importance. Much of the work is independent of this assumption.

In Section 2, as a preliminary, the problem of determining the field generated in a guide of arbitrary section by an assigned tangential electric field in the wall of the guide is solved. In Section 3, some general considerations relating to the slot problem are given, and the analogy of a slot to an antenna is pointed out. The contents of this section are probably not essentially new, but are given here for subsequent convenience.

In Section 4 the problem of a slot in a rectangular guide which transmits only the H_{01}-wave is considered. A quantity called the "voltage" at any point in the slot is introduced, which is the analog of the current in an antenna. This voltage varies approximately sinusoidally along the slot, as does the current in a half-wave antenna. Expressions for the amplitudes of the H_{01}-waves scattered in either direction, in terms of the voltage amplitude, are given, with particular cases, in Eqs. (25–31). It is then shown that, in a certain sense, the slot is equivalent to a network in a transmission line and can be characterized by the reflection and transmission

* This paper is based on Radio Reports Nos. 12 and 13 of the Special Committee on Applied Mathematics of the National Research Council of Canada.

[1] W. H. Watson, *The Physical Principles of Wave Guide Transmission* (Oxford University Press, London, 1947). The writer's work is there referred to, but the notation is different.

[1a] The referee has suggested that the finite conductivity of the walls might lead to large errors. A rough calculation made in the appendix indicates that this is not so, however.

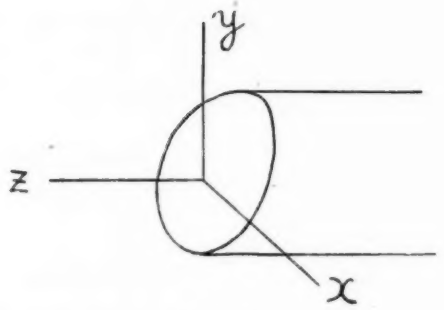


FIG. 1.

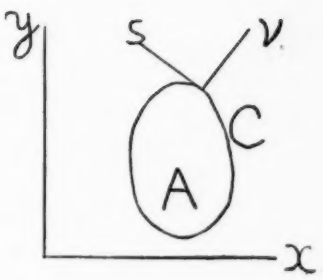


FIG. 2.

coefficients for waves incident from either direction. These coefficients satisfy two identical relations (Eq. (34)), one of which is required for the transmission line analogy to be complete and the other of which imposes a restriction on the type of network to which the slot is equivalent.

In Section 5 it is shown that the transmission coefficients can in part be calculated quite simply from energy considerations, and that this fact suffices to give expressions for the "conductance" or "resistance" of a slot when it is equivalent to a series or shunt element in a transmission line. Detailed formulae for the various relevant cases are given in Eqs. (45–48).

In Section 6, the complete solution of the problem of determining the voltage amplitude in a slot in terms of the amplitude of the incident wave is given (Eqs. (63–68)). This completes the discussion of the problem of a single slot in an infinite guide coupled to free space.

In Section 7, the question of guide-to-guide coupling is considered briefly, and it is shown that this is easily solved when once the solutions for each guide treated separately have been found.

In Section 8, an array of slots is considered, and a set of linear equations is developed for the determination of the voltage amplitudes in the various slots (Eqs. (71–73)). The coefficients in these equations are not hard to calculate when once the calculations for a single slot have been made. In Section 9, the case of a terminated guide is briefly considered.

Not much comparison of the theory with experiment seems to be possible at the present time and in the present state of the calculations. However, some of the formulae of Section 5 agree fairly well with experiment according to Watson (Chapter 6 of reference 1). Detailed calculations for a slot in the broad face of the guide have been in progress for some time, and it is hoped that these will be published soon.

I am greatly indebted to Professor W. H. Watson for many interesting and helpful discussions of the problems involved. I also wish to thank Mr. J. R. Pounder for assistance in the preparation of this paper.

2. THE FIELD GENERATED IN A WAVE-GUIDE OF ARBITRARY SECTION BY AN ASSIGNED TANGENTIAL ELECTRIC FIELD IN THE WALL OF THE GUIDE

As a preliminary, we consider the problem of finding the field generated in an infinite waveguide by any assigned tangential electric field in the wall of the guide. In this section, we shall suppose that the cross section of the guide is arbitrary, since there is here no point in specializing for a rectangular guide. This problem has been solved by Bethe[2] by quite a different method. The method used here, however, offers certain advantages for the slot problem.

Take a system of right-handed axes with the z axis along the guide (Fig. 1), and let A denote the cross section of the guide and C the boundary of A (Fig. 2). Also let ν denote the outward normal to C and s the direction of the tangent, a rotation from ν to s being in the same sense as a rotation from x to y. A suffix ν or s will be used to denote the components of a vector in these directions.

The field within the guide can be expressed in terms of two functions ψ, Ψ, by means of the

[2] H. A. Bethe (not yet published).

formulae:

$$E_x=\frac{\partial^2\psi}{\partial z\partial x}+ik\frac{\partial\Psi}{\partial y},$$

$$E_y=\frac{\partial^2\psi}{\partial z\partial y}-ik\frac{\partial\Psi}{\partial x},$$

$$E_z=\frac{\partial^2\psi}{\partial z^2}+k^2\psi,$$

$$H_x=-ik\frac{\partial\psi}{\partial y}+\frac{\partial^2\Psi}{\partial z\partial x}, \qquad (1)$$

$$H_y=ik\frac{\partial\psi}{\partial x}+\frac{\partial^2\Psi}{\partial z\partial y},$$

$$H_z=\frac{\partial^2\Psi}{\partial z^2}+k^2\Psi,$$

where, in the usual notation, $k=\omega/c=2\pi/\lambda$ (λ=vacuum wave-length), and ψ, Ψ satisfy the wave equations

$$(\nabla^2+k^2)\psi=0, \quad (\nabla^2+k^2)\Psi=0. \qquad (2)$$

We may conveniently refer to ψ, Ψ, as the *generating functions* for E- and H-waves, respectively.

It is assumed that all complex field quantities vary with the time according to the factor $e^{-i\omega t}$, this factor being omitted, and that the actual physical field quantities are the real parts of the corresponding complex expressions.

Let S denote the surface of the guide. Then we regard E_z, E_s as being assigned functions of position on S, it being assumed in the first place that these functions are continuous and possess continuous derivatives. We also assume that E_z, E_s tend to zero sufficiently rapidly as $z\to\pm\infty$. The boundary conditions of the problem are then that E_z, E_s are assigned on S, and that at $z=\pm\infty$ only outgoing and damped waves are present.

Now, inside the guide, E_z satisfies a wave equation similar to (2). Hence, by a well-known formula in the theory of Green's functions

$$E_z(P)=\int_S\frac{\partial G_1(P,P')}{\partial\nu'}E_z(P')dS', \qquad (3)$$

where $P(x, y, z)$ denotes the point at which E_z is estimated and $P'(x', y', z')$ a point in the domain of integration S. The Green's function G_1, for any two points P, P' within S, may be defined as follows:

(1) G_1, regarded as a function of P, satisfies the wave equation $(\nabla^2+k^2)G_1=0$ everywhere inside S except at P'.

(2) $G_1=0$ when P is on S.

(3) As $P\to P'$, G_1 becomes infinite like $1/4\pi r$ where r denotes the distance from P to P'.

Further, from the Maxwell equations, we easily find that, on S,

$$\frac{\partial H_z}{\partial\nu}=\frac{i}{k}\left[\left(\frac{\partial^2}{\partial z^2}+k^2\right)E_s-\frac{\partial^2E_z}{\partial z\partial s}\right].$$

Hence, similarly to (3),

$$H_z(P)=\frac{i}{k}\int_S G_2(P,P')\left[\left(\frac{\partial^2}{\partial z'^2}+k^2\right)E_s(P')-\frac{\partial^2}{\partial z'\partial s'}E_z(P')\right]dS', \qquad (4)$$

where the Green's function G_2 satisfies conditions which are the same as for G_1, except that condition (2) is replaced by $\partial G_2/\partial\nu=0$ on S, and that, in condition (3), $1/4\pi r$ is to be replaced by $-1/4\pi r$.

By repeated integration by parts with respect to z' or s', and by using the conditions that E_z, E_s vanish at $z=\pm\infty$, we can replace (4) by

$$H_z(P)=\frac{i}{k}\int_S\left[\left(\frac{\partial^2}{\partial z'^2}+k^2\right)G_2(P,P')\cdot E_s(P')-\frac{\partial^2G_2(P,P')}{\partial z'\partial s'}E_z(P')\right]dS'. \qquad (5)$$

Equations (3) and (5) now give the field components E_z, H_z everywhere in the guide if G_1, G_2 can be found. Further, the rates of change of E_z, E_s on S can now be made as large as desired, and (3) and (5) will also apply to the case of an aperture in a perfectly conducting guide, when there will, in general, be discontinuities in E_z, E_s at the edge of the aperture.

We shall now outline a method for finding the Green's function G_1, as follows: consider the problem of finding a function $f(x, y, z)$ which

satisfies the equation

$$(\nabla^2+k^2)f=\phi(x, y, z), \tag{6}$$

with the boundary conditions that $f=0$ on S, and that f gives outgoing or damped waves at $z=\pm\infty$. It can then be seen without difficulty that by making

$$\phi(x, y, z)\rightarrow-\delta(x-x')\delta(y-y')\delta(z-z'), \tag{7}$$

where δ denotes Dirac's "delta-function," we obtain the required function G_1.

We suppose f expanded in terms of the functions for E-waves,

$$f=\sum_n a_n(z)\psi_n(x, y), \tag{8}$$

where ψ_n, μ_n $(n=1, 2, \cdots)$ are the normalized eigenfunctions and eigenvalues of the problem:

$$\left.\begin{aligned}\frac{\partial^2\psi_n}{\partial x^2}+\frac{\partial^2\psi_n}{\partial y^2}+\mu_n^2\psi_n=0 \text{ in } A,\\ \psi_n=0 \text{ on } C,\\ \int_A\psi_n^2dxdy=1.\end{aligned}\right\} \tag{9}$$

Substituting (8) in (6), using (9) and the orthogonality relations of the ψ_n, we find

$$(d^2a_n/dz^2)+u_n^2a_n=\int_A\phi(x, y, z)\psi_n(x, y)dxdy, \tag{10}$$

where

$$u_n=(k^2-\mu_n^2)^{\frac{1}{2}}, \tag{11}$$

the positive square root being taken in (11) if u_n is real and the positive-imaginary root if u_n is pure imaginary. We now solve (10) by the method of variation of constants, using the boundary conditions that $a_n(z)\sim e^{iu_nz}$ as $z\rightarrow+\infty$ and $a_n(z)\sim e^{-iu_nz}$ as $z\rightarrow-\infty$ to determine the arbitrary constants. Applying (7), we then finally obtain

$$G_1(P, P')=-\sum_n\frac{1}{2iu_n}\psi_n(x, y)\psi_n(x', y')e^{iu_n|z-z'|}. \tag{12}$$

The function G_2 is found in a similar manner to be[3]

$$G_2(P, P')=\sum_n\frac{1}{2iU_n}\Psi_n(x, y)\Psi_n(x', y')e^{iU_n|z-z'|}, \tag{13}$$

where, analogously to (9), (11), we define Ψ_n, M_n, U_n, the corresponding quantities for H-waves, by:

$$\left.\begin{aligned}\frac{\partial^2\Psi_n}{\partial x^2}+\frac{\partial^2\Psi_n}{\partial y^2}+M_n^2\Psi_n=0 \text{ in } A,\\ \frac{\partial\Psi_n}{\partial\nu}=0 \text{ on } C,\\ \int_A\Psi_n^2dxdy=1,\end{aligned}\right\} \tag{14}$$

$$U_n=(k^2-M_n^2)^{\frac{1}{2}}, \tag{15}$$

the same convention as to the sign of the square root in (15) being used as in (11).

Substituting (12), (13), in (3), (5), we have E_z, H_z at any point in the guide. If we now use (2), and the third and sixth of (1), we obtain (no additive functions evidently being necessary):

$$\psi(P)=-\sum_n\frac{1}{2iu_n\mu_n^2}\psi_n(x, y)\times\int_S\frac{\partial\psi_n(x', y')}{\partial\nu'}e^{iu_n|z-z'|}E_z(P')dS', \tag{16}$$

$$\Psi(P)=\sum_n\frac{1}{2kU_n}\Psi_n(x, y)\times\int_S\Psi_n(x', y')e^{iU_n|z-z'|}E_z(P')dS'\pm\frac{i}{k}\sum_n\frac{1}{2M_n^2}\Psi_n(x, y)\times\int_S\frac{\partial\Psi_n(x', y')}{\partial s'}e^{iU_n|z-z'|}E_z(P')dS', \tag{17}$$

the $+$ or $-$ sign being taken in (17) according as $z>z'$ or $z<z'$.

[3] We omit a term in G_2 which is proportional to $e^{ik|z-z'|}$, since it is evident from (5) that this does not contribute to the field.

Equations (16), (17) are our final results and give the generating functions for E- and H-waves in terms of the assigned values of E_z, E_s on the wall of the guide and the functions defining the free propagation of E- and H-waves in the guide. In the case of a perfectly conducting guide with an aperture, the integrations need, of course, only be extended over the aperture.

3. THE SLOT PROBLEM IN GENERAL AND THE ANTENNA ANALOGY

To solve the slot problem, it is necessary to find fields inside and outside the guide such that the tangential components of **E** and **H** are continuous as we cross the slot, there being a given incident wave in the guide.

Consider for a minute the case where, instead of the wave-guide, we have an infinite perfectly conducting plane with a narrow slot in it and a wave incident on it from one side. From the symmetry (or near-symmetry) of Maxwell's equations in **E** and **H**, it is evident that this problem is essentially the same as that of an antenna in the form of a narrow strip of metal scattering an incident wave, only with the **E**- and **H**-vectors interchanged.[4] Using this analogy, we can see that in the slot problem the only tangential components of **E** and **H** in the slot which are of importance are the component of **H** along the slot and the component of **E** across the slot. Further, near resonance (length of slot approximately equal to an odd number of half-wavelengths), the "voltage," or integral of **E** across the slot, which corresponds to the current in the antenna problem, is approximately sinusoidal with the free-space wave-length, vanishing at the ends of the slot. The complex amplitude of the voltage will depend on the exact length and width of the slot and the data for the incident wave. Off resonance, the voltage is comparatively small.

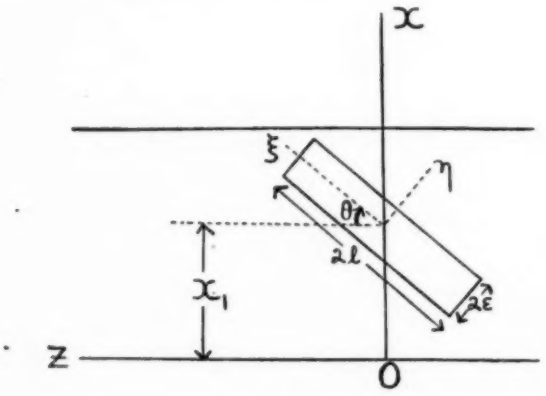


FIG. 3.

A consideration of the way in which this result is obtained will now show that it will still hold approximately for a slot in a wave-guide coupled to free space, or for a slot in guide-to-guide coupling: near resonance the voltage is approximately sinusoidal and vanishes at the ends of the slot, the only exceptions being in the immediate neighborhood of a sharp bend, as when a slot extends around an edge in a rectangular guide. In accordance with the fourth of our fundamental assumptions (Section 1), we shall consider only the case where we are near the first resonance (length of slot $\sim\lambda/2$).

It follows from what has been said above and from the third of our fundamental assumptions that the radiation pattern outside the guide in front of the face containing the slot should be the same as for a half-wave dipole, with the **E**- and **H**-vectors interchanged. Experiments of Watson and collaborators[1] indicate that this conclusion is not well satisfied in some cases. The explanation is that the third assumption referred to above is at fault.

4. A SLOT IN A RECTANGULAR GUIDE AND THE TRANSMISSION LINE ANALOGY

We shall now consider the special case of a rectangular guide which transmits only the H_{01}-wave, one of whose faces contains a slot of length $2l$ and width 2ϵ (Fig. 3). We shall now specify our axes of Section 2 more precisely by taking either direction along one of the edges of the guide lying in the face containing the slot as the z axis, with the x axis through the center of the slot and the y axis along the normal to the guide-face drawn *into* the guide, so that Fig. 3 views the guide-face containing the slot from the *outside*. We shall then specify the position of the slot by x_1, the distance of its center from the z axis, and the angle θ which the direction of the slot makes with the z axis, θ being positive

[4] See Babinet's Principle as formulated by H. G. Booker (not yet published) and E. T. Copson, Proc. Roy. Soc. A186, 100 (1946).

if measured in the same sense as a rotation from Oz to Ox (Fig. 3).

We also take axes (ξ, η) through the center of the slot as shown, the ξ axis being along the slot, and a rotation from the ξ axis to the η axis being in the same sense as a rotation from Oz to Ox. We then define the "voltage" as

$$V(\xi) = \int E_\eta(\xi, \eta) d\eta \tag{18}$$

taken across the slot.

Let a, $b(a>b)$ be the dimensions of the guide (Fig. 4). Suppose now that we have an incident H_{01}-wave traveling in the positive z-direction in an infinite guide. If the slot is in the broad face of the guide, then, with the axes of Fig. 3, the generating function for the incident wave is

$$\Psi^{(0)} = A \cos(\pi x/a) e^{iUz}, \tag{19}$$

where

$$U = (k^2 - \pi^2/a^2)^{\frac{1}{2}} = 2\pi/\lambda_{\text{guide}}, \tag{20}$$

and A is an arbitrary constant, which we shall term the *amplitude of the wave* (at the center of the slot). The slot will scatter H_{01}-waves (as well as damped waves) backwards $(z<0)$ and forwards $(z>0)$. Let these waves have amplitudes B, C, respectively, at the center of the slot, so that they are given by generating functions

$$\begin{aligned} \Psi^{(-)} &= B \cos(\pi x/a) e^{-iUz}, \\ \Psi^{(+)} &= C \cos(\pi x/a) e^{iUz}. \end{aligned} \tag{21}$$

We shall then define the reflection and transmission coefficients α, β by

$$\alpha = B/A, \quad \beta = 1 + C/A. \tag{22}$$

Similarly, if we have an incident H_{01}-wave of amplitude A' traveling in the *negative* z-direction, and waves scattered backwards $(z>0)$ and forwards $(z<0)$ of amplitudes B', C', respectively, we define the reflection and transmission coefficients α', β' by

$$\alpha' = B'/A', \quad \beta' = 1 + C'/A'. \tag{23}$$

If the slot is in the narrow face of the guide, and we keep the axes of Fig. 2, we must write y in place of x in (19) and (21); otherwise the same definitions (22), (23) hold for the reflection and transmission coefficients.

According to the previous section (and as will be verified in detail in Section 6), the voltage in the slot, defined by (18), will be approximately sinusoidal, and vanishing at the ends of the slot, for the case under consideration. Suppose, then, that for an incident H_{01}-wave of amplitude A traveling in the positive direction the voltage is[5]

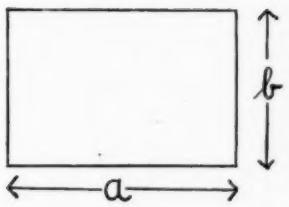


FIG. 4.

$$V = P \cos k\xi, \tag{24}$$

where the constant P may be termed the *voltage amplitude*. Similarly, let the voltage amplitude be P' when an incident wave of amplitude A' is traveling in the negative z-direction.

Formula (17) now gives the amplitudes B, C, of the scattered waves, as defined by (21), which are generated by the voltage (24) in the slot (or the corresponding amplitudes B', C' when the voltage amplitude is P'). Only a single term of the infinite series in (17) is, of course, needed for this, the other terms giving damped waves. In making the calculation we can neglect the width of the slot when once the voltage has been introduced by (18). Using the eigenfunction and eigenvalue for the H_{01}-wave, and making use of (1), it is found that the amplitudes of the scattered waves are, in all cases, given by:

$$\begin{aligned} B/P = C'/P' &= \zeta/\pi^2 kUb, \\ C/P = B'/P' &= \zeta^*/\pi^2 kUb, \end{aligned} \tag{25}$$

where the star denotes the conjugate complex, and ζ is a dimensionless quantity given by

$$\zeta = \int_{-\lambda/4}^{\lambda/4} f(\xi) \cos k\xi \, d\xi, \tag{26}$$

where $f(\xi)/a$ denotes the component of $\mathbf{H}$ along the slot at the point ξ in an H_{01}-wave of *unit* amplitude traveling in the positive z-direction.[6]

[5] This voltage does not vanish at the ends unless the length of the slot is *exactly* $\lambda/2$. It is permissible to suppose this, however, as far as this part of the calculation is concerned (see Section 6).

[6] That ζ must be expressible in this manner becomes clear when the problem is treated by Bethe's method (reference 2).

With this interpretation of $f(\xi)$ the results (25), (26) hold whether the slot is in the broad or the narrow face of the guide. Explicitly, we have, for the slot in the broad face,

$$f(\xi)=e^{iU\xi\cos\theta}\left[\frac{\pi^2}{a}\cos\theta\cos\frac{\pi}{a}(x_1+\xi\sin\theta)\right.$$
$$\left.-i\pi U\sin\theta\sin\frac{\pi}{a}(x_1+\xi\sin\theta)\right],$$

and for the slot in the narrow face,

$$f(\xi)=(\pi^2/a)\cos\theta e^{iU\xi\cos\theta}.$$

We thus find, for the slot in the broad face:

$$\zeta=\frac{\pi^2}{ka}\cos\theta\left[\cos\frac{\pi x_1}{a}I(\theta)-i\sin\frac{\pi x_1}{a}J(\theta)\right]$$
$$+\frac{\pi U}{k}\sin\theta\left[\cos\frac{\pi x_1}{a}J(\theta)-i\sin\frac{\pi x_1}{a}I(\theta)\right], \quad (27)$$

where

$$I(\theta)=\frac{\cos(p\pi/2)}{1-p^2}+\frac{\cos(q\pi/2)}{1-q^2},$$

$$J(\theta)=\frac{\cos(p\pi/2)}{1-p^2}-\frac{\cos(q\pi/2)}{1-q^2},$$

$$p=(U/k)\cos\theta-(\pi/ka)\sin\theta,$$

$$q=(U/k)\cos\theta+(\pi/ka)\sin\theta.$$

As particular cases of this,[7] we have for the *centered inclined slot* $(x_1=a/2)$:

$$\zeta=(i\pi/k)[U\sin\theta I(\theta)+(\pi/a)\cos\theta J(\theta)]; \quad (28)$$

for the *longitudinal slot* $(\theta=0)$:

$$\zeta=2ka\cos(\pi U/2k)\cos(\pi x_1/a); \quad (29)$$

and for the *transverse slot* $(\theta=\pi/2)$:

$$\zeta=-(2\pi ik/U)\cos(\pi^2/2ka)\sin(\pi x_1/a). \quad (30)$$

For the slot in the narrow face we find

$$\zeta=\frac{2\pi^2k}{a(k^2-U^2\cos^2\theta)}\cos\theta\cos(Ul\cos\theta), \quad (31)$$

the result being in this case independent of x_1.

[7] The terminology here used is that of Watson (reference 1).

Formula (25) thus gives the amplitudes of the scattered waves in terms of the voltage amplitudes P or P'. To find the reflection and transmission coefficients (22), (23), we must find the voltage amplitude in terms of the amplitude of the incident wave, i.e., we must find P/A, P'/A'. This is a complicated problem which will be taken up in Section 6. We shall, however, anticipate the results of that section by quoting the formulae for the voltage amplitudes:

$$P/A=\zeta/Ka, \quad P'/A'=\zeta^*/Ka, \quad (32)$$

where ζ denotes the same quantity as defined in (26) above, and K is another (complex) dimensionless constant whose complete expression is given in Section 6. These formulae hold whether the slot is in the broad or the narrow face.

From Eqs. (22), (23), (25), and (32), we now have for the reflection and transmission coefficients:

$$\alpha=\gamma\zeta^2/K, \quad \alpha'=\gamma\zeta^{*2}/K,$$
$$\beta=\beta'=1+\gamma|\zeta|^2/K, \quad (33)$$

where

$$\gamma=1/\pi^2kUab. \quad (33')$$

We thus see that the four coefficients $\alpha, \beta, \alpha', \beta'$ are not independent. They are subject to the *two* relations:

$$\beta=\beta', \quad \alpha\alpha'=(1-\beta)^2. \quad (34)$$

The first of the relations (34) shows that, as far as H_{01}-waves are concerned, the slot is completely analogous to a network connecting two portions of an infinite transmission line, if we define the reflection and transmission coefficients by means (say) of the voltage waves. The equality of the transmission coefficients for waves incident from either direction in a transmission line is, in fact, easily shown to hold under all circumstances by means of ordinary line theory.[8] The second of the relations (34) shows, however, that the equivalent network for a slot is not of the most general type: there is an identical relation connecting the three quantities which characterize the network.

It is now clear that the two coefficients α, β

[8] This was pointed out to the writer by J. R. Pounder. That *some* identical relation must exist between the four coefficients $\alpha, \alpha', \beta, \beta'$ for a transmission line follows from the fact that a network can be characterized by only three independent constants.

(and the associated coefficients α', β' deduced from them by (34)) characterize the slot completely as far as standing-wave-ratio measurements are concerned; and that such quantities as the impedance presented by a slot with a given termination (the terminal impedance being defined by means of a reflection coefficient at the termination), or the equivalent network of a slot, can be worked out in terms of these two coefficients by using ordinary transmission line theory. The only proviso is that the place at which the measurements are made, and the termination, must be at a sufficient distance from the slot for the damped waves to be negligible.

5. CALCULATION OF THE REAL PART OF K. RESISTANCE OR CONDUCTANCE OF SERIES AND SHUNT SLOTS

Although the complete calculation of the constant K in (32) is a complicated matter, which will be considered in the next section, the *real* part of K can be calculated quite simply from energy considerations, as follows.[9] Using the notation of the previous section, and considering the case where the incident wave is traveling in the positive z-direction, the mean flow of energy down the guide (in the z-direction) at $z=-\infty$ is easily calculated to be (in Gaussian units)

$$(\pi\omega bU/16a)(|A|^2-|B|^2). \tag{35}$$

Similarly, the mean flow of energy down the guide at $z=+\infty$ is found to be

$$(\pi\omega bU/16a)[|A|^2+|C|^2+2\Re(AC^*)], \tag{36}$$

where $\Re$ means "real part of." The difference between (35) and (36) must be equal to the mean flow of energy out of the slot. Using our third fundamental assumption (Section 1), this can be calculated in the same way as the flux of energy from a half-wave dipole:[4] it is equal to

$$(73/480\pi^2)c|P|^2. \tag{37}$$

Equating (37) to the difference between (35) and (36), and dividing by $|C|^2$, we obtain, on using (25),

$$\Re\left(\frac{1}{\beta-1}\right)=-1-\frac{73a}{60\pi^3kUb}\left|\frac{P}{C}\right|^2.$$

Hence, from (33), (33′), and (25),

$$\Re(K)=-(73/60\pi)-\gamma|\zeta|^2, \tag{38}$$

$$\Re\left(\frac{1}{\beta-1}\right)=-1-\frac{73}{60\pi\gamma|\zeta|^2}. \tag{39}$$

The results (38), (39) hold, of course, whether the slot is in the broad or the narrow face.

If the equivalent network of the slot can be replaced by simple series or shunt elements, the result (39) suffices to calculate the resistance or conductance of these elements. In terms of the reflection and transmission coefficients, the conditions for the network to be equivalent to a series element are, from ordinary transmission line theory,

$$\alpha=\alpha'=1-\beta, \tag{40}$$

and the admittance of the series element is then

$$Y=(1/2\alpha)-\tfrac{1}{2}. \tag{41}$$

The conditions for the network to be equivalent to a shunt element are

$$\alpha=\alpha'=\beta-1, \tag{42}$$

and the impedance of the shunt element is then

$$Z=-(1/2\alpha)-\tfrac{1}{2}. \tag{43}$$

Reference to (33) then shows that the slot is equivalent to a series element in a transmission line if ζ is a pure imaginary, and to a shunt element if ζ is real. From (27) and (31) we then see that the transverse slot and the centered inclined slot in the broad face are equivalent to series elements, while the longitudinal slot in the broad face and the slot in any position in the narrow face are equivalent to shunt elements.

From (39)–(43) we now have for the "conductance" of the "series" slot or the "resistance" of the "shunt" slot the same expression, namely,

$$G_{\text{series}}=R_{\text{shunt}}=73/120\pi\gamma|\zeta|^2. \tag{44}$$

From (28)–(31), we therefore have:

Transverse slot in broad face (series):

$$\frac{1}{G}=\frac{480\pi^3}{73}\cdot\frac{\gamma k^2}{U^2}\cdot\cos^2\frac{\pi^2}{2ka}\cdot\sin^2\frac{\pi x_1}{a}. \tag{45}$$

[9] We could, of course, obtain the real part from the complete expression for K to be given in Section 6; but the present method is much simpler and forms a useful check on the work.

Centered inclined slot in broad face (series):

$$\frac{1}{G}=\frac{120\pi^3}{73}\cdot\frac{\gamma}{k^2a^2}[Ua\sin\theta I(\theta)+\pi\cos\theta J(\theta)]^2. \quad (46)$$

Longitudinal slot in broad face (shunt):

$$\frac{1}{R}=\frac{480\pi}{73}\cdot\gamma k^2a^2\cdot\cos^2\frac{\pi U}{2k}\cdot\cos^2\frac{\pi x_1}{a}. \quad (47)$$

Slot in narrow face (shunt):

$$\frac{1}{R}=\frac{480\pi^5}{73}\cdot\frac{\gamma}{k^2a^2}\cdot\left[\frac{\cos\theta\cos(Ul\cos\theta)}{1-(U/l)^2\cos^2\theta}\right]^2. \quad (48)$$

In the Eqs. (45)–(48), the half-length of the slot, l, can be put exactly equal to $\lambda/4$, but they should hold approximately for any length of slot close to $\lambda/2$, since the real part of K does not vary rapidly with l. If we define "resonance" by the condition: reactance $=0$, i.e., imaginary part of $K=0$, then actually *at* resonance (i.e., length of slot so that reactance $=0$), we can of course put $R=1/G$. Off resonance, however, we cannot do this, since the imaginary part of K, and hence the reactance or susceptance, varies rapidly with l.

6. CALCULATION OF VOLTAGE AMPLITUDE

We proceed now to the more difficult problem of calculating the voltage amplitude when the incident H_{01}-wave is given.

Referring to Fig. 3, our problem is, in accordance with what was said in Section 3, to find the field component E_η in the slot (or the voltage V) in order that H_ξ may be continuous as we go from inside to outside the slot. We also impose the conditions[10]

$$\left.\begin{aligned} &E_\eta=0 \quad \text{when} \quad \xi=\pm l,\\ \text{i.e.,}\quad &V(\xi)=0 \quad \text{when} \quad \xi=\pm l.\end{aligned}\right\} \quad (49)$$

From (1), (16), (17), we have for the component H_ξ in the slot of the field generated inside the guide by the field E_η:

$$\begin{aligned} H_\xi(\xi,\eta)&=(H_z\cos\theta+H_x\sin\theta)_{y=0}\\ &=\frac{\cos^2\theta}{2k}\sum_n\frac{M_n^2}{U_n}\Psi_n\\ &\quad\times\int_S e^{iU_n|z-z'|}\Psi_n'E_\eta'd\xi'd\eta'\\ &\quad\mp\frac{i}{2k}\cos\theta\sin\theta\sum_n\Psi_n\\ &\quad\times\int_S e^{iU_n|z-z'|}\frac{\partial\Psi_n'}{\partial x'}E_\eta'd\xi'd\eta'\\ &\quad+\frac{k}{2}\sin^2\theta\sum_n\frac{1}{u_n\mu_n^2}\frac{\partial\psi_n}{\partial y}\\ &\quad\times\int_S e^{iu_n|z-z'|}\frac{\partial\psi_n'}{\partial y'}E_\eta'd\xi'd\eta'\\ &\quad\pm\frac{i}{2k}\cos\theta\sin\theta\sum_n\frac{\partial\Psi_n}{\partial x}\\ &\quad\times\int_S e^{iU_n|z-z'|}\Psi_n'E_\eta'd\xi'd\eta'\\ &\quad+\frac{\sin^2\theta}{2k}\sum_n\frac{U_n}{M_n^2}\frac{\partial\Psi_n}{\partial x}\\ &\quad\times\int_S e^{iU_n|z-z'|}\frac{\partial\Psi_n'}{\partial x'}E_\eta'd\xi'd\eta', \end{aligned} \quad (50)$$

where S now denotes the area of the slot. In (50) we have written for brevity

$$\psi_n=\psi_n(x,y),\quad \Psi_n=\Psi_n(x,y),\quad \psi_n'=\psi_n(x',y'),$$
$$\Psi_n'=\Psi_n(x',y'),\quad E_\eta'=E_\eta(\xi',\eta'),$$

and we are to put, after differentiation, $y=y'=0$. The functions ψ_n, Ψ_n and constants μ_n, M_n are the normalized eigenfunctions and eigenvalues for the E- and H-waves in the rectangular guide, as defined by (9) and (14). All these are, of course, *damped* waves except the H_{01}-wave.

Now we have (Fig. 3),

$$\frac{\partial}{\partial\xi}=\cos\theta\frac{\partial}{\partial z}+\sin\theta\frac{\partial}{\partial x},$$

$$\frac{\partial^2}{\partial\xi^2}=\cos^2\theta\frac{\partial^2}{\partial z^2}+\sin^2\theta\frac{\partial^2}{\partial x^2}+2\cos\theta\sin\theta\frac{\partial^2}{\partial z\partial x},$$

[10] These are special cases of the condition that in an aperture in a perfectly conducting screen the component of **E** parallel to the edge of the aperture tends to zero as we tend to the edge from inside the aperture.

so that we can rewrite (50) as follows:

$$H_\xi(\xi,\eta)=\frac{i}{k}\left(\frac{\partial^2}{\partial\xi^2}+k^2\right)\times\int_S G_2(\xi,\eta;\ \xi',\eta')E_\eta'd\xi'd\eta'$$
$$+\frac{\sin^2\theta}{2k}\sum_n\int_S\left(\frac{U_n}{M_n^2}\frac{\partial\Psi_n}{\partial x}\frac{\partial\Psi_n'}{\partial x'}-\frac{1}{U_n}\frac{\partial^2\Psi_n}{\partial x^2}\Psi_n'\right)e^{iU_n|z-z'|}E_\eta'd\xi'd\eta'$$
$$+\frac{k}{2}\sin^2\theta\sum_n\int_S\left(\frac{1}{u_n\mu_n^2}\frac{\partial\psi_n}{\partial y}\frac{\partial\psi_n'}{\partial y'}e^{iu_n|z-z'|}-\frac{1}{U_n}\Psi_n\Psi_n'e^{iU_n|z-z'|}\right)E_\eta'd\xi'd\eta'$$
$$\mp\frac{i}{2k}\cos\theta\sin\theta\sum_n\int_S\left(\Psi_n\frac{\partial\Psi_n'}{\partial x'}+\Psi_n'\frac{\partial\Psi_n}{\partial x}\right)\times e^{iU_n|z-z'|}E_\eta'd\xi'd\eta', \quad (51)$$

where G_2 is the Green's function given by (13). In (51) we are to put $y=y'=0$ after differentiation, and $G_2(\xi,\eta;\xi',\eta')$ means that we are to do the same with G_2.

The integral occurring in the first term in (51) tends to infinity as $\epsilon\to 0$ on account of the singularity in the Green's function G_2, but the remaining three terms remain finite as $\epsilon\to 0$.[11] Making $\epsilon\to 0$ in the last three terms of (51), therefore, and introducing the voltage $V(\xi)$ defined by (18), we can, with an error of the order of ϵ/l, replace (51) by

$$H_\xi(\xi,\eta)=\frac{i}{k}\left(\frac{\partial^2}{\partial\xi^2}+k^2\right)\times\int_S G_2(\xi,\eta;\xi',\eta')E_\eta(\xi',\eta')d\xi'd\eta'$$
$$+\int_{-l}^{l}F(\xi,\xi')V(\xi')d\xi', \quad (52)$$

where

$$F(\xi,\xi')=\frac{\sin^2\theta}{2k}\sum_n\left(\frac{U_n}{M_n^2}\frac{\partial\Psi_n}{\partial x}\frac{\partial\Psi_n'}{\partial x'}-\frac{1}{U_n}\frac{\partial^2\Psi_n}{\partial x^2}\Psi_n'\right)e^{iU_n|z-z'|}$$
$$+\frac{k}{2}\sin^2\theta\sum_n\left(\frac{1}{u_n\mu_n^2}\frac{\partial\psi_n}{\partial y}\frac{\partial\psi_n'}{\partial y'}e^{iu_n|z-z'|}-\frac{1}{U_n}\Psi_n\Psi_n'e^{iU_n|z-z'|}\right)\mp\frac{i}{2k}\cos\theta\sin\theta$$
$$\times\sum_n\left(\Psi_n\frac{\partial\Psi_n'}{\partial x'}+\Psi_n'\frac{\partial\Psi_n}{\partial x}\right)e^{iU_n|z-z'|}, \quad (53)$$

and in (53) we are to put, after differentiation,

$$x=x_1+\xi\sin\theta,\quad z=\xi\cos\theta,\quad x'=x_1+\xi'\sin\theta,\quad z'=\xi'\cos\theta,\quad y=y'=0. \quad (54)$$

We now consider the field generated *outside* the guide. With our third fundamental assumption (Section 1), this is simply the field due to E_η across the slot regarded as being in an infinite conducting plane. This problem can be solved by introducing the "associate Hertz vector," or it can be inferred by analogy with the corresponding problem of a thin antenna in the form of a flat strip having a given current distribution.[4] We omit the details and give only the final result:

$$H_\xi(\xi,\eta)=\frac{i}{2\pi k}\left(\frac{\partial^2}{\partial\xi^2}+k^2\right)\times\int_S\frac{e^{ikr}}{r}E_\eta(\xi',\eta')d\xi'd\eta', \quad (55)$$

where

$$r=[(\xi-\xi')^2+(\eta-\eta')^2]^{\frac{1}{2}}.$$

Equations (52) and (55) now give the component H_ξ in the slot of the fields generated inside and outside the guide, respectively, by E_η. According to the continuity condition mentioned at the beginning of this section, the difference between (55) and (52) must be equal to the component H_ξ of the incident field. We denote this latter component by $Af(\xi)/a$, where A is the amplitude of the incident wave at the center of the slot, so that $f(\xi)$ is identical with the function introduced in (26). The continuity condition then

[11] Each of the three terms actually consists of the sum of two parts, the infinities of which cancel. The considerations of Section 3 show that the only singularity which should occur in H_ξ as $\epsilon\to 0$ is correctly given by the first term of (51).

gives

$$\frac{i}{k}\left(\frac{d^2}{d\xi^2}+k^2\right)\int_S\left[G_2(\xi,\eta;\xi',\eta')-\frac{1}{2\pi}\frac{e^{ikr}}{r}\right]E_\eta(\xi',\eta')d\xi'd\eta'$$

$$=-\int_{-l}^{l}F(\xi,\xi')V(\xi')d\xi'-Af(\xi)/a. \quad (56)$$

We have written an ordinary, instead of a partial, derivative on the left-hand side of (56), since the right-hand side is a function of ξ only, and the same must therefore be true of the left-hand side. (This is consistent with the assumption —see Section 3—that we can neglect the component of **H** across the slot.)

Regarding (56) as a differential equation for the left-hand side, and solving by the method of variation of constants, the general solution can be written

$$\frac{i}{k}\int_S\left[G_2(\xi,\eta;\xi',\eta')-\frac{1}{2\pi}\frac{e^{ikr}}{r}\right]E_\eta(\xi',\eta')d\xi'd\eta'$$

$$=-\frac{i}{k}\cos k\xi\int_{-l}^{\xi}d\xi\int_{-l}^{l}d\xi'\cdot\sin k\xi F(\xi,\xi')V(\xi')$$

$$-\frac{i}{k}\sin k\xi\int_{-l}^{\xi}d\xi\int_{-l}^{l}d\xi'\cdot\cos k\xi F(\xi,\xi')V(\xi')$$

$$+\frac{A}{ka}\cos k\xi\int_{-l}^{\xi}\sin k\xi f(\xi)d\xi$$

$$-\frac{A}{ka}\sin k\xi\int_{-l}^{\xi}\cos k\xi f(\xi)d\xi$$

$$+C_1\cos k\xi+C_2\sin k\xi, \quad (57)$$

where C_1, C_2 are arbitrary constants. Equation (57) is now an integral equation for the determination of E_η, or $V(\xi)$, the constants C_1, C_2 being determined by the boundary conditions (49). We shall adopt a method of solution analogous to one used by Hallén[12] in connection with a similar integral equation occurring in an antenna problem. We rewrite the integral on the left-hand side of (57) as

$$\int_S\left(G_2-\frac{1}{2\pi}\frac{e^{ikr}}{r}\right)E_\eta(\xi,\eta')d\xi'd\eta'$$

$$+\int_S\left(G_2-\frac{1}{2\pi}\frac{e^{ikr}}{r}\right)$$

$$\times[E_\eta(\xi',\eta')-E_\eta(\xi,\eta')]d\xi'd\eta'. \quad (58)$$

The contribution to the first integral in (58) from a strip of width $d\eta'$ parallel to the η axis and at a distance η' from it is

$$E_\eta(\xi,\eta')d\eta'\int_{-l}^{l}\left(G_2-\frac{1}{2\pi}\frac{e^{ikr}}{r}\right)d\xi'. \quad (59)$$

Now from the definition of G_2 it follows that, allowing for the "image effect" when the point (x, y, z) is near a wall of the guide, $G_2\sim-1/2\pi r$ when r is small. Hence, since the major contribution to the integral in (59) comes from points where r is small, the value of the integral is approximately

$$-(1/4\pi)\int_{-l}^{l}d\xi'/r.$$

This latter integral has, with an error of order ϵ/l, the value $2\log(2l/|\eta-\eta'|)$ except when ξ is near $\pm l$. But since E_η vanishes when $\xi=\pm l$, this exception is unimportant. We may, therefore, with an error of order ϵ/l, replace the first integral in (58) by

$$-\frac{1}{2\pi}\int_{-\epsilon}^{\epsilon}\log\frac{2l}{|\eta-\eta'|}E_\eta(\xi,\eta')d\eta'. \quad (60)$$

The second integral in (58), on the other hand, remains finite as $\epsilon\to0$, and may (with an error of order ϵ/l) be replaced by

$$\int_{-l}^{l}\left[G_2(\xi,\xi')-\frac{1}{2\pi}\frac{e^{ik|\xi-\xi'|}}{|\xi-\xi'|}\right][V(\xi')-V(\xi)]d\xi',$$

where $G_2(\xi,\xi')$ means that we are to make $\epsilon\to0$ in $G_2(\xi,\eta;\xi',\eta')$.

We now see that (with our approximations) the integral in (60) must be independent of η, which shows that E_η must vary across the slot in the same way that the electrostatic charge density varies across an infinite conducting strip of the

[12] E. Hallén, Nova Acta Reg. Soc. Upsaliensis **11**, No. 4 (1938).

same width. We can therefore write

$$\int_{-\epsilon}^{\epsilon} \log\frac{2l}{|\eta-\eta'|} E_\eta(\xi, \eta')d\eta' = \log\frac{2l}{\tilde{\epsilon}} V(\xi), \quad (61)$$

where the constant $\tilde{\epsilon}$, defined by (61), can be calculated from the known solution of the electrostatic problem of the infinite conducting strip. Obviously $\tilde{\epsilon}$ is of the order of ϵ and it really makes no difference, to the order to which we are working, whether we take $\tilde{\epsilon}=\epsilon$ or use the correct value. As however, the calculation is easily made, we shall use the correct value; it proves to be $\tilde{\epsilon}=\epsilon/2$. Collecting our results, we now see that the integral equation (57) can be replaced by

$$-\frac{i}{2\pi k}\log\frac{4l}{\epsilon}V(\xi) = \frac{A}{k}\cos k\xi\int_{-l}^{\xi}\sin k\xi f(\xi)d\xi - \frac{A}{k}\sin k\xi\int_{-l}^{\xi}\cos k\xi f(\xi)d\xi + C_1\cos k\xi + C_2\sin k\xi + \Lambda(\xi), \quad (62)$$

where

$$\Lambda(\xi) = -\frac{i}{k}\int_{-l}^{l} G_2(\xi, \xi')[V(\xi')-V(\xi)]d\xi' + \frac{i}{2\pi k}\int_{-l}^{l}\frac{e^{ik|\xi-\xi'|}}{|\xi-\xi'|}[V(\xi')-V(\xi)]d\xi' + \frac{i}{k}\cos k\xi\int_{-l}^{\xi}d\xi\int_{-l}^{l}d\xi'\sin k\xi F(\xi, \xi')V(\xi') - \frac{i}{k}\sin k\xi\int_{-l}^{\xi}d\xi\int_{-l}^{l}d\xi'\cos k\xi F(\xi, \xi')V(\xi').$$

We can now solve (62) by successive approximations, regarding $\log 4l/\epsilon$ as being large, in accordance with our second fundamental assumption. For a first approximation, we neglect $\Lambda(\xi)$ in (62); for a second approximation, we substitute the first approximation for $V(\xi)$ in $\Lambda(\xi)$, and so on. We then find $V(\xi)$ as a series in the small quantity $(\log 4l/\epsilon)^{-1}$—or more accurately in $(2\log 4l/\epsilon)^{-1}$. The arbitrary constants C_1, C_2 are then determined by using the boundary conditions (49). We are here only concerned with the case $l\sim\lambda/4$, i.e. $kl\sim\pi/2$. We then find that we must proceed to at least the *second* approximation to find the constant C_1, but that having found it the term proportional to C_1 in the first approximation is the major term. Confining ourselves to this approximation, the solution is as follows:

$$V(\xi) = P\cos k\xi,$$

where

$$P/A = \zeta/Ka, \quad (63)$$

ζ being given by (26), and where

$$K = -\frac{i}{\pi}\log\frac{4l}{\epsilon}\cos kl - i\int_{-l}^{l}[G_2(l, \xi')+G_2(-l, \xi')]\cos k\xi'd\xi' + \frac{i}{\pi}\int_{-l}^{l}\frac{e^{ik(l-\xi')}}{l-\xi'}\cos k\xi'd\xi' - \int_{-l}^{l}\int F(\xi, \xi')\cos k\xi\cos k\xi'd\xi d\xi'. \quad (64)$$

In all terms in K except the first, we are to put $l=\lambda/4$, $kl=\pi/2$.

If the incident wave is traveling in the negative z-direction, and is of amplitude A', we find similarly that the corresponding voltage amplitude P' is given by

$$P'/A' = \zeta^*/Ka. \quad (65)$$

The results (63) and (65) are those which have already been used in Section 4.

If we write

$$K = K_1 + iK_2,$$

then K_1 is a comparatively simple expression which has already been given in (38). For K_2 we can write, to the order to which we are working,

$$K_2 = (\delta L/\lambda)\log(2\lambda/w) + K_2', \quad (66)$$

where $w=2\epsilon$ is the width of the slot, $\delta L = 2l-\lambda/2$ is the *excess* of the length of the slot over a half-wave-length, and K_2' denotes the imaginary part of the expression for K in (64) omitting the first term, so that K_2' does not depend (to our order of approximation) on the length or width of the slot. From (13) and (53), it will be seen that K_2' is given by quite a complicated expression in the form of a doubly-infinite series.

If the slot is in the broad face of the guide, we have explicitly from (13) and (53), on using the

eigenfunctions and eigenvalues for the rectangular guide:[13]

$$G_2(\xi,\xi')=\frac{2}{iab}\sum_{n,m}\frac{\delta_{nm}}{u_{nm}}\cos\frac{n\pi x}{a}\times\cos\frac{n\pi x'}{a}e^{iu_{nm}|\xi-\xi'|\cos\theta}, \quad (67)$$

$$F(\xi,\xi')=\frac{2\sin^2\theta}{kab}\sum_{n,m}\delta_{nm}\left(\frac{n\pi}{a}\right)^2\times\left[\frac{u_{nm}}{\mu_{nm}^2}\sin\frac{n\pi x}{a}\sin\frac{n\pi x'}{a}+\frac{1}{u_{nm}}\cos\frac{n\pi x}{a}\cos\frac{n\pi x'}{a}\right]e^{iu_{nm}|\xi-\xi'|\cos\theta}+\frac{2k\sin^2\theta}{ab}\sum_{n,m}\delta_{nm}\left[\frac{(m\pi/b)^2}{u_{nm}\mu_{nm}^2}\sin\frac{n\pi x}{a}\sin\frac{n\pi x'}{a}-\frac{1}{u_{nm}}\cos\frac{n\pi x}{a}\cos\frac{n\pi x'}{a}\right]e^{iu_{nm}|\xi-\xi'|\cos\theta}, \quad (68)$$

where

$$x=x_1+\xi\sin\theta, \qquad x'=x_1+\xi'\sin\theta,$$
$$\mu_{nm}^2=(n\pi/a)^2+(m\pi/b)^2, \quad u_{nm}=(k^2-\mu_{nm}^2)^{\frac{1}{2}},$$
$$\delta_{nm}=1, \text{ unless } n \text{ or } m=0,$$
$$=\tfrac{1}{2}, \quad \text{if } n \text{ or } m=0.$$

The summations in (67) and (68) are taken with respect to n and m from 0 to ∞ with the exception of the values $n=m=0$. The positive, or positive-imaginary, square root for u_{nm} is to be taken.

If the slot is in the narrow face of the guide, we need merely interchange a and b in (67) and (68), and, of course, use the appropriate expression for ζ (i.e., (31) instead of (27)).

7. GUIDE-TO-GUIDE COUPLING

Consider now the case where the slot couples two infinite guides—guide 1, where we have an incident (exciting) wave, and guide 2, where waves are generated. Suppose that Fig. 3 gives the orientation of the slot relative to the standard axes for guide 1. Then the orientation of the slot relative to the standard axes for guide 2 will be given by an exactly similar figure (with, of course, different parameters x_1, θ in general). The only difference is that, if we keep the direction of the ξ axis the same in both guides and preserve the relative orientations of the (ξ, η) axes, then the η axes are in opposite directions for the two guides, so that the voltage is the *negative* of that defined by (18) when considered from the point of view of guide 2. This is because the normal to the guide-face drawn *into* the guide has opposite directions for the two guides.

We can then carry through an exactly similar analysis to that used in Section 6, the only difference being that instead of the solution *outside* the guide there used, we must use the solution in guide 2. Obviously the voltage is still approximately sinusoidal and we find for the amplitude P, if the incident wave is of amplitude A and traveling in the positive z-direction,

$$P/A=\zeta_1/K_{12}a_1, \quad (69)$$

where the subscripts 1 in ζ, a refer to guide 1, and where

$$K_{12}=-\frac{i}{\pi}\log\frac{4l}{\epsilon}\cos kl-i\int_{-l}^{l}[G_2^{(1)}(l,\xi')+G_2^{(1)}(-l,\xi')]\cos k\xi'\,d\xi'-i\int_{-l}^{l}[G_2^{(2)}(l,\xi')+G_2^{(2)}(-l,\xi')]\cos k\xi'\,d\xi'-\int_{-l}^{l}\int F^{(1)}(\xi,\xi')\cos k\xi\cos k\xi'\,d\xi d\xi'-\int_{-l}^{l}\int F^{(2)}(\xi,\xi')\cos k\xi\cos k\xi'\,d\xi d\xi', \quad (70)$$

the superscripts 1, 2 in the functions G_2, F in (70) referring to guides 1, 2, respectively. If the incident wave is traveling in the negative z-direction, we must write ζ_1^* in place of ζ_1 in (69).

Energy considerations similar to those used in Section 5 show that

$$\Re(K_{12})=-\gamma_1|\zeta_1|^2-\gamma_2|\zeta_2|^2,$$

[13] The term proportional to $\cos\theta\sin\theta$ in (53) has been omitted from $F(\xi,\xi')$, since it can be seen that this vanishes on integration.

where γ_1, γ_2 are the constants γ defined by (33) for the two guides.

Knowing the voltage amplitude, the reflection and transmission coefficients for guide 1 and the field generated in guide 2 are, of course, easily calculated. It will be seen that if the calculations have been made for the slot in each guide separately, the calculations for the guide-to-guide coupling are readily performed. In fact, reference to (64) and (70) shows that most of the terms in K_{12} are simply the sum of the corresponding terms for the K's of the two guides. It will also be observed that our third fundamental assumption (Section 1) is not used in this case, so that the results should be more accurate.

8. SLOT ARRAYS

We now consider the case of an array of N slots numbered 1, 2, $\cdots$, N in the same face of an infinite guide, and coupled to empty space. It is evident that if the lengths of all the slots are approximately $\lambda/2$ the voltage in each slot is still approximately sinusoidal and the problem is to calculate the amplitudes $P_1, P_2, \cdots, P_N$ in terms of the amplitude of the incident wave. Having done this, we can find the radiation pattern outside the guide, and the reflection and transmission coefficients, etc., inside the guide.

Considering, say, the jth slot, we proceed as before, but now in finding the component of **H** along the slot we must include, in addition, the contributions from the voltages in all the other slots, when reckoning the field either inside or outside the guide. We must also allow for the fact that the incident wave has a different phase at the center of each slot. The calculations go very much the same as before, finding a first approximation, and then a second approximation, to the voltage in each slot. We omit the details and quote the final result: the voltage amplitudes are determined by the set of linear equations:

$$\sum_{j'=1}^{N} p_{jj'}P_{j'} = (A/a)\zeta_j e^{iUz_j}, \quad j' = 1, \cdots, N, \quad (71)$$

where it has been assumed that the incident wave is traveling in the positive z-direction and has amplitude A at some *arbitrary* point in the guide, z_j being the distance down the guide from this point, measured in the positive z-direction, of the center of the jth slot, and where

$$p_{jj} = K_j, \quad (72)$$

$$p_{jj'} = \int_{-l}^{l} [F_{jj'}^{(\text{out})}(\xi) - F_{jj'}^{(\text{in})}(\xi)] \cos k\xi d\xi, \quad j \neq j'. \quad (73)$$

In (71), (72), ζ_j, K_j mean the constants ζ, K, already defined, calculated for the jth slot, while in (73), $F_{jj'}^{(\text{out})}(\xi)$ denotes the component of **H** along the jth slot due to a sinusoidal voltage of *unit* amplitude in the j'th slot, calculated for *outside* the guide, and $F_{jj'}^{(\text{in})}(\xi)$ has a similar meaning when calculated for *inside* the guide. $F_{jj'}^{(\text{in})}$ can be found from formulae already given; it would probably be sufficient to confine ourselves to the contribution from the H_{01}-wave generated by the jth slot, so that $F_{jj'}^{(\text{in})}(\xi)$ would be quite a simple expression. $F_{jj'}^{(\text{out})}$ can be found by the method outlined in Section 6 for finding the field outside the guide, and again yields a simple closed expression.

There is thus no particular difficulty about calculating the coefficients $p_{jj'}(j \neq j')$, while the coefficients p_{jj} are given by the calculations for a *single* slot. The coefficients $p_{jj'}(j \neq j')$ express, of course, the interaction between the slots. For distant slots, $F_{jj'}^{(\text{out})}$ decreases as $(z_j - z_{j'})^{-1}$, while $F_{jj'}^{(\text{in})}$ remains constant. While the calculation of all the coefficients $p_{jj'}$ and the solution of the set of equations (71) would in general present a formidable task, it would appear that in the arrays of practical importance considerable simplifications are possible. The K_j might be determined experimentally.

If the incident wave is traveling in the negative z-direction we need merely write $\zeta_j^* e^{-iUz_j}$ in place of ζe^{iUz_j} in (71).

9. THE CASE OF A TERMINATED GUIDE

So far we have dealt entirely with infinite guides. In practice, however, the guide may have an unmatched termination, so that we must take account of the termination (a guide with a matched termination can, of course, be treated as an infinite guide). We shall indicate briefly here how a termination of arbitrary impedance can be dealt with, and it will appear that this case can be deduced very simply from that of the infinite guide already dealt with.

Suppose we take the z axis pointing towards the termination, and that we have available solutions for the infinite guide for incident waves traveling in either direction. Assuming that all slots are at a sufficient distance from the termination (and in the negative z-direction from it) for all waves but the H_{01}-wave to be damped out at the termination, the infinite-guide-solution for a wave incident in the positive z-direction will, in the neighborhood of the termination, be an H_{01}-wave whose generating function is, say, $\chi_1^{(+)}$. Similarly, the infinite-guide-solution for a wave incident in the negative z-direction will, in the neighborhood of the termination, be a superposition of H_{01}-waves traveling in both directions, so that its generating function will be of the form $\chi_2^{(-)}+\alpha\chi_2^{(+)}$, where α is a constant of the nature of a reflection coefficient (the superscripts $+$, $-$ refer to waves traveling in the positive and negative z-directions, respectively).

We now consider the field given by

$$\chi_1^{(+)}+\lambda[\chi_2^{(-)}+\alpha\chi_2^{(+)}],$$

where λ is an arbitrary constant. By proper choice of λ we can now make the ratio of the amplitudes of the incident to the reflected waves at the termination anything we please, and hence can give the termination an arbitrary "impedance." Thus a linear combination of the two solutions for the infinite guide, for incident waves traveling in opposite directions, gives the solution for the semi-infinite guide with an arbitrary termination. In particular, the voltage amplitude in any slot will be

$$P_1+\lambda P_2$$

where P_1, P_2, are the voltage amplitudes for the two solutions for the infinite guide.

APPENDIX

On the Correction Due to Imperfect Conductivity of the Guide Walls

To take account completely of the finite conductivity of the walls of the guide would be quite possible, but would lead to elaborate computations. The following very rough calculation, in which we consider the problem of Section 5, indicates, however, that the error caused by the assumption of perfect conductivity is not likely to be great.

The chief losses caused by the imperfect conductivity of the walls will occur near the slot, where the fields are large. In this region we can estimate the tangential components of the magnetic field by analogy with the problem of the half-wave antenna (see Section 3). We thus find that, near the slot, the major component of $\mathbf{H}$ is that perpendicular to the length of the slot, and is of amount

$$H=\frac{P}{2\pi\eta}\left[(l-\xi)\frac{e^{ikr_2}}{r_2}-(l+\xi)\frac{e^{ikr_1}}{r_1}\right],$$

where (ξ, η) are the coordinates of a point in the guide face containing the slot relative to the axes of Fig. 3, r_1, r_2 are the distances of the point from the ends of the slot, and the notation is otherwise that used before.[14] Except near the ends of the slot, this can be replaced approximately, if η is small, by

$$H=\frac{P}{2\pi\eta}[e^{ik(l-\xi)}-e^{ik(l+\xi)}]. \tag{A1}$$

If, as we may assume, the depth of penetration is small compared with the thickness of the wall, we have approximately for the tangential *electric* field when imperfect conductivity is allowed for,[15]

$$\mathbf{E}=(1-i)\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}(\mathbf{H}\times\mathbf{n}), \tag{A2}$$

where $\mathbf{n}$ is a unit vector drawn into the guide wall, f is the frequency, and σ the conductivity of the guide wall.

Using (A1), (A2), we thus have for the mean flux of energy into the guide wall, per unit area,

$$\frac{c}{8\pi}\Re(\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{n}=\frac{c}{8\pi}\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}|H|^2$$
$$=\frac{c}{16\pi^3}\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}\frac{|P|^2}{\eta^2}(1-\cos 2k\xi).$$

The total mean flux of energy into the guide walls is thus of the order

$$2\cdot\frac{c}{16\pi^3}\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}|P|^2\cdot l'\int_\epsilon\frac{d\eta}{\eta^2}\sim\frac{c}{8\pi^3}\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}|P|^2\frac{l'}{\epsilon}, \tag{A3}$$

where l' is a length of the order of the length of the slot, and 2ϵ is, as before, the width of the slot.

Comparing (A3) with the expression (37) for the mean flux of energy out of the slot, we see that the error involved in assuming perfect conductivity of the walls is of the order

$$\frac{1}{\pi}\left(\frac{f}{4\sigma}\right)^{\frac{1}{2}}\frac{l'}{\epsilon}.$$

For Copper and a wave-length of the order of 1 cm., $f:4\sigma$ is of the order 10^{-6}. A reasonable value of l'/ϵ would be about 30. The above error is then about 10^{-2} or one percent.

[14] See, for instance, Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 457, formula 76, on interchanging $\mathbf{E}$ and $\mathbf{H}$ and allowing for the difference in units. Note that Stratton's I_0 corresponds to $(c/2\pi)P$ in the present case. We have disregarded an unimportant phase factor.

[15] See, for instance, reference 14, p. 534, formula 47, allowing for the difference in units.

Scattering of Electromagnetic Radiation by a Thin Circular Ring in a Circular Wave Guide

PAULA FEUER AND EDWARD S. AKELEY
Department of Physics, Purdue University, Lafayette, Indiana
(Received April 18, 1947)

The scattering of a $TE_{1,1}$ wave by a thin, perfectly conducting circular ring suspended in a circular wave guide with perfectly conducting walls is considered, and approximate expressions are obtained for the scattering cross section, resonance maximum, and half-width as a function of the width of the ring.

I. INTRODUCTION

IT has been observed[1] that obstacles of certain special sizes and shapes, when placed in a wave guide, reflect practically all radiation of a particular frequency incident upon them. Radiation of frequency different from this resonance frequency is partly reflected and partly transmitted. An obstacle of this type is the resonant ring—a conducting ring of metal placed perpendicular to the wave-guide axis. Resonance is found to occur when the circumference of the ring is somewhat greater than the wave-length in free space.

This paper concerns itself with the theory of such a ring.[2] In particular, theoretical expressions will be obtained for the field that exists in a circular wave guide with perfectly conducting walls when a $TE_{1,1}$ wave is scattered by an infinitely thin, perfectly conducting, narrow circular ring suspended in the guide. The ratio of the scattered to the incident field at large distances from the ring will be examined in the neighborhood of resonance and the resonance maximum and half-width will be determined approximately, as a function of the width of the ring.

The electromagnetic field that exists within a circular wave guide is expressible as a linear combination of the fields known as the Transverse Electric and Transverse Magnetic Fields which are the solutions of the Maxwell Equations in cylindrical coordinates.

[1] Sperry Gyroscope Company, "Microwave transmission design data" (Publication No. 23–80), p. 117.

[2] The problem considered here has been treated by another method in an unpublished paper by Dr. J. Schwinger.

"Transverse Electric" Field Distribution[3]

$$\begin{aligned} E_r &= i\mu\omega(\partial\psi/\partial\phi), & H_r &= \pm ih(\partial\psi/\partial r), \\ E_\phi &= -i\mu\omega(\partial\psi/\partial r), & H_\phi &= -(ih/r)(\partial\psi/\partial\phi), \\ E_z &= 0, & H_z &= (k^2-h^2)\psi. \end{aligned} \tag{1}$$

"Transverse Magnetic" Field Distribution[3]

$$\begin{aligned} E_r &= \pm ih(\partial\psi/\partial r), & H_r &= -(ik^2/\mu\omega r)(\partial\psi/\partial\phi), \\ E_\phi &= \pm(ih/r)(\partial\psi/\partial\phi), & H_\phi &= (ik^2/\mu\omega)(\partial\psi/\partial r), \\ E_z &= (k^2-h^2)\psi, & H_z &= 0, \end{aligned} \tag{2}$$

where

$$\psi = e^{in\phi}J_n(\lambda r)e^{\pm ihz-i\omega t}, \tag{3}$$

or

$$\psi = e^{in\phi}H_n(\lambda r)e^{\pm ihz-i\omega t}. \tag{4}$$

J_n and H_n are Bessel functions of the first kind and $\lambda=(k^2-h^2)^{\frac{1}{2}}$ where k is 2π divided by the wave-length in free space. The real part of h is 2π divided by the wave-length in the guide.

Let R represent the radius of the wave guide. When ψ is of the form (3) and h is such that $(k^2-h^2)^{\frac{1}{2}}R$ equals the mth root of the nth Bessel function J_n, then E_ϕ, E_z, and H_r of (2) vanish at $r=R$, and the field (2) becomes the $TM_{n,m}$ mode. Similarly when h is such that $(k^2-h^2)^{\frac{1}{2}}R$ equals the mth root of the first derivative of the nth Bessel function, the field (1) becomes the $TE_{n,m}$ mode.

It is assumed that k and R are such that the $TE_{1,1}$ mode only is propagated, other modes being exponentially damped. The $TE_{1,1}$ wave, incident upon the ring (see Fig. 1), sets up

[3] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), first edition, p. 360.

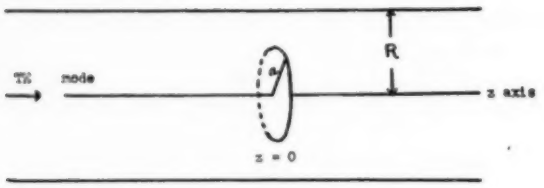


FIG. 1. Resonant ring in the circular wave guide.
Note: TE should be $TE_{1,1}$.

currents in it which in turn alter the currents on the walls of the guide. For, since the total tangential components (E_ϕ and E_z) of the electric field and the normal component (H_r) of the magnetic field must vanish on the perfectly conducting walls, additional currents must appear on the walls to counteract the field of the currents in the ring. The resultant field at any point in the guide is thus produced by (a) the incident $TE_{1,1}$ wave, (b) the currents in the ring, and (c) the additional currents on the walls of the guide caused by the presence of the ring.

II. THE FIELD OF THE CURRENTS IN THE RING

Figure 2 is a diagram of the ring in the plane $z=0$. It will be assumed that the current density set up in the ring is

$$J_r=J_z=0, \quad J_\theta=\frac{C\cos\theta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}.$$

$2b$ is the width of the ring, η the distance from the circular axis of the ring, and C a constant. The proportionality to $\cos\theta$ is suggested because the θ component of the electric field of the $TE_{1,1}$ wave is proportional to $\cos\theta$ (θ is measured from the line on which E_θ is maximum). C will later be determined in such a way as to make the total tangential electric field vanish on the ring. The radial distribution factor $1/[\pi b(1-\eta^2/b^2)^{\frac{1}{2}}]$ is the same as that which would exist in an infinitely long, infinitely thin, perfectly conducting narrow strip in space when an electro-

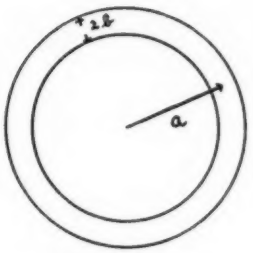


FIG. 2. Resonant ring in the plane $z=0$.

magnetic wave fell upon it. Used here, it is a good approximation provided that b is sufficiently small compared to the wave-length.

The field produced by the ring currents at a point $P(r, \phi, z)$ in the wave guide is evaluated from the relations

$$V=\frac{a}{4\pi\epsilon}\int_{-b}^{b}d\eta\int_{0}^{2\pi}d\theta\frac{e^{ikd-i\omega t}}{d}\rho,$$

$$\mathbf{A}=\frac{\mu a}{4\pi}\int_{-b}^{b}d\eta\int_{0}^{2\pi}d\theta\frac{e^{ikd-i\omega t}}{d}\mathbf{J},$$

$$H=\frac{\nabla\times\mathbf{A}}{\mu},$$

$$E=-\nabla V-\frac{\partial\mathbf{A}}{\partial t}. \tag{5}$$

V and $\mathbf{A}$ are, respectively, the scalar and vector potentials, d is the distance from the current element $ad\theta$ to the point P, and $\mathbf{J}$ and ρ (determined from $\nabla\cdot\mathbf{J}=-\partial\rho/\partial t$) the current density and charge density, respectively. It is convenient to express e^{ikd}/d in terms of the cylinder functions as follows:

$$\frac{e^{ikd}}{d}=\frac{i}{2}\int_{-\infty}^{+\infty}dh e^{-ihz}\sum_{-\infty}^{+\infty}J_m(\lambda a)H_m(\lambda r)e^{im(\theta-\phi)}$$

where $\lambda=(k^2-h^2)^{\frac{1}{2}}$. The electromagnetic field of the ring currents is thus found to be:

$$E_r=C\sin\phi\int_{-b}^{b}\frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}dh e^{-ihz}\left[\frac{\lambda J_1(\lambda a+\lambda\eta)H_1'(\lambda r)}{4\epsilon\omega}+\frac{\mu\omega(a+\eta)}{8}\right.$$
$$\left.\times\left(J_0(\lambda a+\lambda\eta)H_0(\lambda r)-J_2(\lambda a+\lambda\eta)H_2(\lambda r)\right)\right],$$

$$E_\phi = C\cos\phi \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh e^{-ihz}\left[\frac{J_1(\lambda a+\lambda\eta)H_1(\lambda r)}{4\epsilon\omega r} + \frac{\mu\omega(a+\eta)}{8} \times\{J_0(\lambda a+\lambda\eta)H_0(\lambda r)+J_2(\lambda a+\lambda\eta)H_2(\lambda r)\}\right], \quad (6)$$

$$E_z = -\frac{Ci\sin\phi}{4\epsilon\omega} \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh h e^{-ihz} J_1(\lambda a+\lambda\eta)H_1(\lambda r),$$

$$H_r = -\frac{C\cos\phi}{8} \int_{-b}^{b} \frac{d\eta(a+\eta)}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh h e^{-ihz}[J_0(\lambda a+\lambda\eta)H_0(\lambda r)+J_2(\lambda a+\lambda\eta)H_2(\lambda r)],$$

$$H_\phi = \frac{C\sin\phi}{8} \int_{-b}^{b} \frac{d\eta(a+\eta)}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh h e^{-ihz}[J_0(\lambda a+\lambda\eta)H_0(\lambda r)-J_2(\lambda a+\lambda\eta)H_2(\lambda r)],$$

$$H_z = \frac{Ci\cos\phi}{8} \int_{-b}^{b} \frac{d\eta(a+\eta)}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh e^{-ihz}[J_0(\lambda a+\lambda\eta)H_0'(\lambda r)+J_2(\lambda a+\lambda\eta)H_2'(\lambda r)] + \frac{Ci\cos\phi}{4} \int_{-b}^{b} \frac{d\eta(a+\eta)}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh e^{-ihz} \frac{J_2(\lambda a+\lambda\eta)H_2(\lambda r)}{r}.$$

A prime indicates the derivative of the cylinder function with respect to its argument. The factor $e^{-i\omega t}$ has been omitted from each component.

III. THE FIELD OF THE ADDITIONAL CURRENTS ON THE WAVE-GUIDE WALLS CAUSED BY THE PRESENCE OF THE RING

Since the total E_ϕ, E_z, and H_r must vanish on the walls of the guide, the additional currents set up on the walls must produce field strengths E_ϕ'', E_z'', and H_r'' which at $r=R$ cancel E_ϕ, E_z and H_r. Furthermore, the field of these currents must be expressible as a linear combination of solutions of the form (1) and (2). A field which satisfies these conditions may be constructed as follows. Consider

$$E_z'' = C\int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh(k^2-h^2)Ke^{-ihz}J_1(\lambda r)\sin\phi,$$

where

$$K = \frac{i}{4\epsilon\omega}\frac{h}{k^2-h^2}\frac{J_1(\lambda a+\lambda\eta)H_1(\lambda R)}{J_1(\lambda R)}.$$

E_z'' is the negative of E_z at $r=R$. Moreover, E_z'' has the form of the z component of the electric field of (2) with $n=1$. Associated with E_z'' (as indicated by (2)) are

$$\begin{aligned}
E_r'' &= -C\sin\phi \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dhihe^{-ihz}K\lambda J_1'(\lambda r),\\
E_\phi'' &= -C\cos\phi \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dhihe^{-ihz}K\frac{J_1(\lambda r)}{r},\\
H_r'' &= -C\cos\phi \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh\frac{ik^2}{\mu\omega}e^{-ihz}K\frac{J_1(\lambda r)}{r},\\
H_\phi'' &= C\sin\phi \int_{-b}^{b} \frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} dh\frac{ik^2}{\mu\omega}e^{-ihz}K\lambda J_1'(\lambda r),\\
H_z'' &= 0.
\end{aligned} \quad (7)$$

It is now convenient to write E_ϕ, given by (6), as

$$E_\phi = -\frac{C\cos\phi}{4\epsilon\omega}\int_{-b}^{b}\frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}dh\frac{h^2}{k^2-h^2}e^{-ihz}J_1(\lambda a+\lambda\eta)\frac{H_1(\lambda r)}{r}$$
$$+\frac{C\cos\phi}{4\epsilon\omega}\int_{-b}^{b}\frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}dh\frac{k^2}{k^2-h^2}e^{-ihz}J_1(\lambda a+\lambda\eta)\frac{H_1(\lambda r)}{r}$$
$$-\frac{C\mu\omega\cos\phi}{8}\int_{-b}^{b}\frac{d\eta(a+\eta)}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}dhe^{-ihz}[J_0(\lambda a+\lambda\eta)H_0(\lambda r)+J_2(\lambda a+\lambda\eta)H_2(\lambda r)].$$

Then, at $r=R$, E_ϕ'' is the negative of the first term of E_ϕ. A field component which will cancel the second and third terms of E_ϕ at $r=R$ is

$$E_\phi''' = -i\mu\omega C\cos\phi\int_{-b}^{b}\frac{d\eta}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}dhe^{-ihz}LJ_1'(\lambda r)$$

where

$$L=\frac{1}{4i\omega^2\epsilon\mu}\frac{k^2}{k^2-h^2}\frac{J_1(\lambda a+\lambda\eta)H_1(\lambda R)}{RJ_1'(\lambda R)}+\frac{(a+\eta)i}{8}\frac{J_0(\lambda a+\lambda\eta)H_0(\lambda R)+J_2(\lambda a+\lambda\eta)H_2(\lambda R)}{J_1'(\lambda R)}.$$

E_ϕ''' has the form of the ϕ component of the electric field of (1) with $n=1$. When the Transverse Electric field which may be associated with E_ϕ''' is added to the Transverse Magnetic field (7), the resultant

$$\begin{aligned}
E_r' &= -Ci\sin\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dhe^{-ihz}[hKJ_1'(\lambda r)+\mu\omega LJ_1(\lambda r)],\\
E_\phi' &= -Ci\cos\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dhe^{-ihz}\left[hK\frac{J_1(\lambda r)}{r}+\mu\omega LJ_1'(\lambda r)\right],\\
E_z' &= C\sin\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dh(k^2-h^2)e^{-ihz}KJ_1(\lambda r),\\
H_r' &= -Ci\cos\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dhe^{-ihz}\left[\frac{k^2}{\mu\omega}K\frac{J_1(\lambda r)}{r}+hLJ_1'(\lambda r)\right],\\
H_\phi' &= Ci\sin\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dhe^{-ihz}\left[\frac{k^2}{\mu\omega}\lambda KJ_1'(\lambda r)+hLJ_1(\lambda r)\right],\\
H_z' &= C\cos\phi\int_{-b}^{b}\frac{d\eta}{\tau}\int_{-\infty}^{+\infty}dh(k^2-h^2)e^{-ihz}LJ_1(\lambda r),
\end{aligned}\tag{8}$$

where

$$\tau=\frac{1}{\pi b(1-\eta^2/b^2)^{\frac{1}{2}}}$$

satisfies the required boundary conditions. The condition that H_r' cancel H_r at $r=R$ is automatically fulfilled. It should be pointed out however that (8) is not unique. Addition to (8) of a TM or TE mode gives a field which still satisfies the boundary conditions. And indeed, in the next section, (8) will be modified so as to satisfy the additional condition that no reflections occur (that no waves originate) at one end of the guide.

IV. THE FIELD FAR FROM THE RING AND THE SCATTERING CROSS SECTION

In order to determine the constant C, $E_\phi+E_\phi'$ will now be evaluated in detail. C will then be determined by making the total ϕ component of the electric field vanish on the circular axis of the ring. $E_\phi+E_\phi'$ (given by (6) and (8)) is evaluated for $z>0$ by a complex integration over h over the path indicated in Fig. 3. The points avoided on the real axis correspond to $h=-(k^2-\beta_1^2/R^2)^{\frac{1}{2}}$ and $h=(k^2-\beta_1^2/R^2)^{\frac{1}{2}}$ where β_1 is the first root of J_1'. An infinite number of singularities occur on the imaginary axis (where $(k^2-h^2)^{\frac{1}{2}}R$ is real) since the denominators $J_1(\lambda R)$ and $J_1'(\lambda R)$ have infinitely many real roots. The integration yields,

$$
\begin{aligned}
E_\phi+E_\phi' = {} & \frac{C\pi(\mu)^{\frac{1}{2}}k\cos\phi}{4(\epsilon)^{\frac{1}{2}}}\int_{-b}^{b}\frac{d\eta}{\tau}\frac{J_1\left(\beta_1\frac{a+\eta}{R}\right)N_1(\beta_1)J_1'\left(\beta_1\frac{r}{R}\right)}{\beta_1(k^2R^2-\beta_1^2)^{\frac{1}{2}}J_1''(\beta_1)} \\
& \times\left\{\exp\left[i\left(k^2-\frac{\beta_1^2}{R^2}\right)^{\frac{1}{2}}z\right]-\exp\left[-i\left(k^2-\frac{\beta_1^2}{R^2}\right)z\right]\right\}-\frac{C\pi(\mu)^{\frac{1}{2}}k\cos\phi}{8(\epsilon)^{\frac{1}{2}}R} \\
& \times\int_{-b}^{b}\frac{d\eta(a+\eta)}{\tau}\frac{\beta_1\left(J_0\left(\beta_1\frac{a+\eta}{R}\right)N_0(\beta_1)+J_2\left(\beta_1\frac{a+\eta}{R}\right)N_2(\beta_1)\right)J_1'\left(\beta_1\frac{r}{R}\right)}{(k^2R^2-\beta_1^2)^{\frac{1}{2}}J_1''(\beta_1)} \\
& \left\{\exp\left[i\left(k^2-\frac{\beta_1^2}{R^2}\right)^{\frac{1}{2}}z\right]-\exp\left[-i\left(k^2-\frac{\beta_1^2}{R^2}\right)^{\frac{1}{2}}z\right\}+\frac{C\pi i(\mu)^{\frac{1}{2}}\cos\phi}{2(\epsilon)^{\frac{1}{2}}krR}\int_{-b}^{b}\frac{d\eta}{\tau}\right. \\
& \times\sum_{m=1}^{\infty}\frac{(\gamma_m^2-k^2R^2)^{\frac{1}{2}}J_1\left(\gamma_m\frac{a+\eta}{R}\right)N_1(\gamma_m)J_1\left(\gamma_m\frac{r}{R}\right)}{\gamma_m J_1'(\gamma_m)}\exp\left[-\left(\frac{\gamma_m^2}{R^2}-k^2\right)^{\frac{1}{2}}z\right] \\
& +\frac{C\pi i(\mu)^{\frac{1}{2}}k\cos\phi}{2(\epsilon)^{\frac{1}{2}}}\int_{-b}^{b}\frac{d\eta}{\tau}\sum_{m=2}^{\infty}\left[\frac{J_1\left(\beta_m\frac{a+\eta}{R}\right)N_1(\beta_m)}{\beta_m(\beta_m^2-k^2R^2)^{\frac{1}{2}}}\right. \\
& \left.\left.-\frac{\beta_m(a+\eta)\left(J_0\left(\beta_m\frac{a+\eta}{R}\right)N_0(\beta_m)+J_2\left(\beta_m\frac{a+\eta}{R}\right)N_2(\beta_m)\right)}{2R(\beta_m^2-k^2R^2)^{\frac{1}{2}}}\right]\right. \\
& \left.\times\frac{J_1'\left(\beta_m\frac{r}{R}\right)}{J_1''(\beta_m)}\exp\left[-\left(\frac{\beta_m^2}{R^2}-k^2\right)^{\frac{1}{2}}z\right]\right. . \qquad (9)
\end{aligned}
$$

γ_m and β_m are the roots of J_1 and J_1', respectively. The relations $\omega=k/(\epsilon\mu)^{\frac{1}{2}}$ and $H_m(x)=J_n(x)+iN_n(x)$ have been used.

Terms containing the factors $\exp\{i[k^2-(\beta_1^2/R^2)]^{\frac{1}{2}}z\}$ and $\exp\{-i[k^2-(\beta_1^2/R^2)]^{\frac{1}{2}}z\}$ represent $TE_{1,1}$ waves traveling to the right and left, respectively. It will now be assumed that there are no sources of waves and no reflections at the right (as is the case when the terminating impedance equals the characteristic impedance of the guide) and terms with the factor $\exp\{-i[k^2-(\beta_1^2/R^2)]^{\frac{1}{2}}z\}$ will be

discarded. This amounts to adding to (8) a $TE_{1,1}$ wave which just cancels the terms corresponding to $TE_{1,1}$ waves traveling to the left. As was indicated in the previous section, the resultant field still satisfies the other boundary conditions. The last two terms of (9) represent the higher modes that are exponentially damped as they leave the ring. The field component (9) together with

$$E_\phi^{(0)} = -i\mu\omega(\beta_1/R)\cos\phi J_1'[\beta_1(r/R)]\exp[i(k^2-\beta_1^2/R^2)^{\frac{1}{2}}z] \tag{10}$$

the ϕ component of the electric field of the incident $TE_{1,1}$ wave, represents the total ϕ component of the electric field in the wave guide.

The constant C is now determined by setting equal to zero the total ϕ component of the electric field (given by (9) with terms having the factor $\exp[-i(k^2-\beta_1^2/R^2)^{\frac{1}{2}}z]$ discarded+(10)) on the circular axis of the ring; that is, at $r=a$, $z=0$.

$$\frac{C\pi(\mu)^{\frac{1}{2}}k\cos\phi}{4(\epsilon)^{\frac{1}{2}}}\int_{-b}^{b}\frac{d\eta}{2b}\left[\frac{J_1[\beta_1(a+\eta/R)]N_1(\beta_1)}{\beta_1(k^2R^2-\beta_1^2)^{\frac{1}{2}}J_1''(\beta_1)}-\frac{a+\eta}{2R}\frac{\beta_1(J_0[\beta_1(a+\eta/R)]N_0(\beta_1)+J_2[\beta_1(a+\eta/R)]N_2(\beta_1)}{(k^2R^2-\beta_1^2)^{\frac{1}{2}}J_1''(\beta_1)}\right]$$

$$\times J_1'\left(\beta_1\frac{a}{R}\right)+\frac{C\pi i(\mu)^{\frac{1}{2}}\cos\phi}{2(\epsilon)^{\frac{1}{2}}kaR}\int_{-b}^{b}\frac{d\eta}{2b}\sum_{m=1}^{\infty}\frac{(\gamma_m^2-k^2R^2)^{\frac{1}{2}}J_1[\gamma_m(a+\eta/R)]N_1(\gamma_m)J_1[\gamma_m(a/R)]}{\gamma_m J_1'(\gamma_m)}$$

$$+\frac{C\pi i(\mu)^{\frac{1}{2}}k\cos\phi}{2(\epsilon)^{\frac{1}{2}}}\int_{-b}^{b}\frac{d\eta}{2b}\left[\sum_{m=2}^{\infty}\frac{J_1[\beta_m(a+\eta/R)]N_1(\beta_m)J_1'[\beta_m(a/R)]}{\beta_m(\beta_m^2-k^2R^2)^{\frac{1}{2}}J_1''(\beta_m)}\right.$$

$$\left.-\frac{a+\eta}{2R}\frac{\beta_m(J_0[\beta_m(a+\eta/R)]N_0(\beta_m)+J_2(\beta_m(a+\eta/R)]N_2(\beta_m))}{(\beta_m^2-k^2R^2)^{\frac{1}{2}}J_1''(\beta_m)}J_1'\left(\beta_m\frac{a}{R}\right)\right]$$

$$-i\mu\omega\frac{\beta_1}{R}\cos\phi J_1'\left(\beta_1\frac{a}{R}\right)+\frac{Ci(\mu)^{\frac{1}{2}}\cos\phi\ln(2/e)}{2\pi(\epsilon)^{\frac{1}{2}}a}\left(\frac{1}{ka}-ka\right)=0. \tag{11}$$

Note that the factor $1/\tau=1/[\pi b(1-\eta^2/b^2)^{\frac{1}{2}}]$ has been replaced by the constant $1/(2b)$ and that an additional term $(Ci(\mu)^{\frac{1}{2}}\cos\phi\ln(2/e)/2\pi(\epsilon)^{\frac{1}{2}}a)[(1/ka)-ka]$ has been inserted. The constant distribution factor $1/(2b)$ was used to simplify the integration over η. The added term corrects for this. It represents the difference in the ϕ component of the electric field calculated with the distribution factor $1/\tau$ and then with the factor $1/(2b)$, and it is evaluated as follows:

Consider the electric field at a point P on the ring. It can be shown that those contributions to the field at P which depend on whether $1/[\pi b(1-\eta^2/b^2)^{\frac{1}{2}}]$ or $1/(2b)$ is used, come from the current and charge in a small region surrounding P. Therefore, (5), the expression for the ϕ component of the electric field due to the currents in the ring is integrated from $-a\epsilon$ to $a\epsilon$ (instead of from 0 to 2π) where $\epsilon>0$ is small. To do this, d is set equal to $(2a(a+\eta)-2a(a+\eta)\cos(\theta-\phi)+\eta^2)^{\frac{1}{2}}$ and the integrands are expanded in powers of η/R with second and higher order terms neglected. This is carried out using both distribution factors, the difference in the two results being the added term in (11).

Equation (11) is simplified (see appendix) and solved for C to yield

$$C=\frac{ik(\beta_1a/R)J_1'[\beta_1(a/R)]}{iS\ln(b/R)+iU-T}$$

where

$$S=\frac{1}{2\pi}\left(\frac{1}{ka}-ka\right),\quad T=\frac{\pi ka^2}{4R}\frac{\beta_1J_1'^2[\beta_1(a/R)]N_1'(\beta_1)}{(k^2R^2-\beta_1^2)^{\frac{1}{2}}J_1''(\beta_1)}$$

and

$$U=-\frac{1}{2\pi ka}\left[\frac{\sin\gamma_{p+1}(b/R)}{\gamma_{p+1}(b/R)}+\int_0^{\gamma_{p+1}b/R}du\frac{1-\cos u}{u}-\ln\Gamma\gamma_{p+1}\right]+\frac{1}{2ka}\sum_{m=p+1}^{\infty}\frac{\sin 2\gamma_m(a/R)\sin\gamma_m(b/R)}{\gamma_m^2(b/R)}$$

$$+\frac{\cos 2\gamma_m(a/R)}{2\gamma_m}\left[\frac{\cos\gamma_m(b/R)}{\gamma_m(a/R)}-\frac{\sin\gamma_m(b/R)}{\gamma_m^2(ab/R^2)}\right]+\frac{ka}{2\pi}\left[\frac{\sin\beta_{q+1}(b/R)}{\beta_{q+1}(b/R)}+\int_0^{\beta_{q+1}b/R}du\frac{1-\cos u}{u}-\ln\Gamma\beta_{q+1}\right]$$

$$+\frac{ka}{2}\sum_{m=q+1}^{\infty}\frac{\sin 2\beta_m(a/R)\sin\beta_m(b/R)}{\beta_m^2(b/R)}+\frac{\cos 2\beta_m(a/R)}{2\beta_m}\left[-\frac{\cos\beta_m(b/R)}{\beta_m(a/R)}+\frac{\sin\beta_m(b/R)}{\beta_m^2(ab/R^2)}\right]$$

$$-\frac{\pi^2 a}{4kaR}\sum_{m=1}^{p}(\gamma_m^2-k^2R^2)^{\frac{1}{2}}J_1^2\left(\gamma_m\frac{a}{R}\right)N_1^2(\gamma_m)$$

$$+\frac{\pi^2 a}{4kaR}\sum_{m=1}^{q}\frac{\beta_m^2 J_1'^2[\beta_m(a/R)]N_1'^2(\beta_m)}{(\beta_m^2-k^2R^2)^{\frac{1}{2}}[1-(1/\beta_m^2)]}+\frac{\ln(2/e)}{2}\left(\frac{1}{ka}-ka\right).$$

Γ is the Euler constant.

It should be emphasized that although the total tangential electric field must vanish everywhere on the ring, this condition is exactly satisfied only for $r=a$. It is satisfied to a good approximation everywhere else on the ring provided the width of the ring is sufficiently small compared with the wave-length.

The transmitted E_ϕ far from the ring is now found by inserting the value for C and letting $z\to\infty$ in (9). The integration over η can then be performed, after expanding the remaining terms in powers of η/R, neglecting terms of second and higher order. The ratio of this transmitted field to the incident field (10) is

$$E_\phi]_{z\to\infty}/E_\phi^{(0)}]_{z\to\infty}=T/[iS\ln(b/R)+iU-T]=Ae^{i\psi}. \tag{12}$$

V. THE RESONANCE MAXIMUM

The value of k for which the amplitude A of the ratio (12) is maximum, is now determined. Consider

$$A^2=T^2/\{[S\ln(b/R)+U]^2+T^2\}. \tag{13}$$

For b/R small, $\ln(b/R)$ is a large negative number, and since $S\ln(b/R)$ is the dominant term in the denominator, resonance is to be expected when S is small. S vanishes for $ka=1$. Hence it is assumed that resonance occurs for $k_{max}a=1+\delta$ where δ is small. Moreover, A^2 is obviously maximum (equal to 1) when

$$S\ln(b/R)=-U. \tag{14}$$

Expanding S and U around $ka=1$, so that

$$\begin{aligned}S&=-\delta/\pi+\delta^2/2\pi+\cdots,\\ U&=U_0+U_1\delta+\cdots,\end{aligned} \tag{15}$$

and solving (14) for δ, the second order approximation

$$\delta=\frac{\pi u_0}{\ln(b/R)}+\frac{\pi^2}{2\ln^2(b/R)}(U_0^2+2U_0U_1)$$

is obtained. Hence

$$k_{max}a=1+\frac{\pi U_0}{\ln(b/R)}+\frac{\pi^2}{2\ln^2(b/R)}\times(U_0^2+2U_0U_1). \tag{16}$$

It can be shown that $U_0<0$ and hence that $[\pi U_0/\ln(b/R)]>0$. Since k is 2π divided by the wave-length in free space, it follows from (16) that resonance occurs for $2\pi a$, the circumference of the ring, greater than the wave-length in free space. As the width of the ring is made smaller, the resonance condition approaches $k_{max}a=1$ or $2\pi a=$ wave-length in free space.

Also, when $S\ln(b/R)$ is set equal to $-U$ in

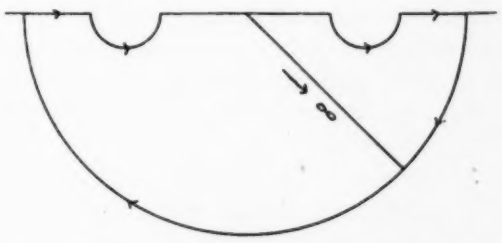


FIG. 3. Path of integration in the complex plane.

(12), it is found that $(E_\phi]_{z\to\infty}/E_\phi^{(0)}]_{z\to\infty})=-1$. Hence, at resonance, the transmitted field is exactly out of phase with the incident field and all the energy is reflected.

VI. THE HALF-WIDTH

From (10) it is seen that A^2 falls to the value $\frac{1}{2}$ when

$$S\ln(b/R)+U=\pm T. \tag{17}$$

Expanding T around $ka=1$ ($T=T_0+T_1\delta+\cdots$), and using (17), it is found that the values of δ for which (17) holds are

$$\delta_{\frac{1}{2}}=\frac{\pi}{\ln(b/R)}(U_0+T_0)+\frac{\pi^2}{\ln^2(b/R)}\left(\frac{(U_0+T_0)^2}{2}+(U_0+T_0)(U_1+T_1)\right)+\cdots,$$

$$\delta_{-\frac{1}{2}}=\frac{\pi}{\ln(b/R)}(U_0-T_0)+\frac{\pi^2}{\ln^2(b/R)}\left(\frac{(U_0-T_0)^2}{2}+(U_0-T_0)(U_1-T_1)\right)+\cdots,$$

giving a half-width of

$$\delta_{\frac{1}{2}}-\delta_{-\frac{1}{2}}=\frac{2\pi T_0}{\ln(b/R)}+\frac{2\pi^2}{\ln^2(b/R)}\times(U_0T_0+U_0T_1+T_0U_1)+\cdots.$$

As the width of the ring decreases, the half-width decreases, i.e. resonance becomes sharper.

The relative half-width is

$$\frac{1}{Q}=\frac{\delta_{\frac{1}{2}}-\delta_{-\frac{1}{2}}}{k_{\max}a}=\frac{2\pi T_0}{\ln(b/R)}+\frac{2\pi^2}{\ln^2(b/R)}(U_0T_1+U_1T_0).$$

It is worth noting that although the ring considered here was infinitely thin (it was assumed to be in the plane $z=0$), the theory of the ring of small finite thickness is only slightly different, the difference being in the correction term added in (11).

It should also be pointed out that the problem of transmission of a $TE_{1,1}$ wave by a circular slit in a partition in a circular wave guide is quite analogous to the problem of the resonant ring. That is, the resonant slit could be treated by assuming a (fictitious) magnetic current to be set up in the slit by the $TE_{1,1}$ wave. The expression for the field of this magnetic current density could be chosen so as to satisfy the necessary boundary conditions; and the fields far from the slit, the resonance maximum and the half-width could then be calculated in the same way as for the resonant ring.

APPENDIX

The first term of (11) is expanded in powers of η/R, with terms of second and higher order neglected, and then integrated over η.

In the second and third terms of (11), asymptotic representations may be used for the Bessel functions when m is large. Consider in particular the second term of (11). Using the relation $J_1'(\gamma_m)=-2/[\pi\gamma_m N_1(\gamma_m)]$ and the asymptotic expressions

$$J_1\left(\gamma_m\frac{a}{R}\right)\sim\left(\frac{R}{\pi\gamma_m a}\right)^{\frac{1}{2}}\left(-\cos\gamma_m\frac{a}{R}+\sin\gamma_m\frac{a}{R}\right),$$

$$N_1^2(\gamma_m)\sim 2/\pi\gamma_m,$$

$$(\gamma_m^2-k^2R^2)^{\frac{1}{2}}\sim\gamma_m,$$

this becomes, after integration over η,

$$-\frac{C\pi^2 i(\mu)^{\frac{1}{2}}}{4(\epsilon)^{\frac{1}{2}}kaR}\sum_{m=1}^{p}(\gamma_m^2-k^2R^2)^{\frac{1}{2}}J_1^2[\gamma_m(a/R)]N_1^2(\gamma_m),$$

$$-\frac{Ci(\mu)^{\frac{1}{2}}}{2(\epsilon)^{\frac{1}{2}}ka^2}\sum_{m=p+1}^{\infty}\frac{\sin\gamma_m(b/R)}{\gamma_m^2(b/R)}, \tag{18}$$

$$+\frac{Ci(\mu)^{\frac{1}{2}}}{2(\epsilon)^{\frac{1}{2}}ka^2}\sum_{m=p+1}^{\infty}\frac{\sin2\gamma_m(a/R)\sin\gamma_m(b/R)}{\gamma_m^2(b/R)}, \tag{19}$$

$$-\frac{Ci(\mu)^{\frac{1}{2}}}{4(\epsilon)^{\frac{1}{2}}ka^3}\sum_{m=p+1}^{\infty}\frac{\cos2\gamma_m(a/R)}{\gamma_m}\times\left(-\frac{\cos\gamma_m(b/R)}{\gamma_m/R}+\frac{\sin\gamma_m(b/R)}{\gamma_m^2(b/R)}\right). \tag{20}$$

The series from 1 to p represents those terms of the infinite series for which m is too small for asymptotic expressions to be used. Since $\gamma_m b/R$ is a slowly varying function, (18) may be re-

placed by the integral

$$-\frac{Ci(\mu)^{\frac{1}{2}}}{2(\epsilon)^{\frac{1}{2}}ka^2\pi}\int_{\gamma_{j+1}b/R}^{\infty}du\frac{\sin u}{u^2}=\frac{Ci(\mu)^{\frac{1}{2}}}{2(\epsilon)^{\frac{1}{2}}ka^2\pi}\left[\ln\Gamma\frac{\gamma_{p+1}b}{R}\right.$$
$$\left.-\int_0^{\gamma_{p+1}b/R}du\frac{1-\cos u}{u}-\frac{\sin\gamma_{p+1}(b/R)}{\gamma_{p+1}(b/R)}\right].$$

Γ is the Euler constant.

The series (19) and (20) are rapidly convergent.

The third term of (11) is similarly evaluated using the relations

$$J_1(\beta_m)=2/[\pi\beta_m N_1'(\beta_m)],$$
$$J_1''(\beta_m)=2/[\pi\beta_m N_1'(\beta_m)][-1+(1/\beta_m^2)],$$

and the asymptotic expressions

$$J_1'[\beta_m(a/R)]\sim J_0[\beta_m(a/R)]\sim J_2[\beta_m(a/R)]$$
$$\sim(R/\pi\beta_m a)^{\frac{1}{2}}[\cos\beta_m(a/R)+\sin\beta_m(a/R)],$$
$$N_1'(\beta_m)\sim N_0(\beta_m)\sim N_2(\beta_m)$$
$$\sim(1/\pi\beta_m)^{\frac{1}{2}}(\sin\beta_m-\cos\beta_m),$$
$$1-(1/\beta_m^2)\sim 1,\quad(\beta_m^2-k^2R^2)^{\frac{1}{2}}\sim\beta_m.$$

A Resonant Cavity Linear Accelerator*

A. B. Cullen, Jr. and J. H. Greig
Rouss Physics Laboratory, University of Virginia, Charlottesville, Virginia
(Received June 16, 1947)

A folded rectangular wave guide cavity, resonant at a frequency of 2800 megacycles, is used to accelerate electrons which are injected at 2000 volts energy to an energy of 300 kilovolts. Three accelerating stages are provided, the power source being a type 4J33 radar magnetron operating on a duty cycle of 2×10^{-4}. Power output of the magnetron is 800 kilowatts, the power consumed by the cavity being approximately 250 kilowatts. Design data for the cavity are obtained from slotted wave guide models of the cavity sections.

THE principle of operation of the linear accelerator is well known.[1] This paper describes such a device in which high accelerating fields are obtained by using a high powered magnetron to excite a resonant cavity. The resonant electron accelerator described employs a folded rectangular wave guide cavity of variable cross section to accelerate 2000-volt electrons to 300 kilovolts. The 250-kilowatt pulse power dissipated by the cavity is supplied by a radar magnetron operating in the 10-cm band. The electrons are accelerated through field maxima of the dominant transverse electric mode in which the cavity is operated. The electron energies so obtained agreed well with the design value, 300 kv. Visual observations of the deflected beam indicate that the energy spread was certainly less than 5 percent.

* This work was done in part under Navy Contract NOrd-7873.

[1] R. Wideroe, Archiv. f. Electrotechnik **21**, 387 (1929); Sloan and Lawrence, Phys. Rev. **38**, 2021 (1931); L. B. Snoddy, H. Trotter, W. Ham, and J. W. Beams, J. Frank. Inst. **223**, 55 (1937).

PRINCIPLE OF OPERATION

A plan of the cavity is shown in Fig. 1. The resonant portion of the cavity extends from the tuning plunger at F to the inductive window at E, with the first, third, and fifth maxima of the electric field lined up along the electron path A-B-C-D. The second and fourth maxima occur in the bends, and are necessary to obtain the proper polarity. The electrons enter through a quarter-inch hole in the center of the cavity wall at A, continuing through similar holes at B, C, and D.

The width AB is such that an electron entering at A, when the field is passing through zero and increasing, will be accelerated towards B and arrive there a half-cycle later. At this time the field between B and C is zero and increasing in such a direction that a second acceleration is produced. Thus the electron receives three accelerations in three half-cycles of the 2800-mc driving power. In order to satisfy this phase sequence it is necessary that the distance between AB and BC be an electrical wave-length.

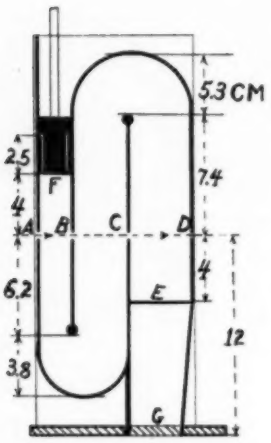


FIG. 1. Plan view of resonant cavity.

CAVITY DESIGN

It was decided to use a cavity of length sufficient to produce three separate accelerations, because these are enough to show up any serious faults with the system. The first dimension to be decided upon was the dimension perpendicular to the plane of Fig. 1, which shall be called a. This was chosen to be equal to the width of standard wave guide, 2.840″. It might be noted that since the power lost in the cavity is a function of this dimension, some other value might give greater efficiency. The power loss in a straight rectangular cavity is given (for the lowest mode) by[2]

$$P=\frac{R_s\lambda^2\epsilon E_0^2}{8\mu}\left[\frac{bd}{a^2}+\frac{ab}{d^2}+\frac{a}{2d}+\frac{d}{2a}\right], \tag{1}$$

where a, b, and d are the dimensions of the cavity. R_s is the surface resistivity of the walls, λ is the free-space wave-length of the exciting power, ϵ and μ are the dielectric constant and permeability of the medium, respectively, and E_0 is the maximum electric field. The dimensions a and d are related by the equation

$$2d=\frac{\lambda}{\left(1-\left(\frac{\lambda}{2a}\right)^2\right)^{\frac{1}{2}}}. \tag{2}$$

It can be seen from Eq. (1) that the power loss is not a very sensitive function of a, hence its value can be chosen for mechanical convenience. The b dimension is measured in the direction of the accelerating field, hence it must be equal to the distance the electrons travel in a half-cycle for a given accelerating field.

In order to determine dimension b it is necessary to compute the distance traveled by an electron in a uniform electric field which varies sinusoidally with time. First the force equation

$$eE_0\sin\omega t=\frac{d}{dt}\left(\frac{m\beta c}{(1-\beta^2)^{\frac{1}{2}}}\right) \tag{3}$$

is integrated, and solving the result for $\beta(1-\beta^2)^{-\frac{1}{2}}$ we obtain

$$\beta(1-\beta^2)^{-\frac{1}{2}}=A+B(1-\cos\omega t), \tag{4}$$

in which $A=\beta_0(1-\beta_0^2)^{-\frac{1}{2}}$, and $B=eE_0/m\omega c$.

The distance travelled by the electron in a half-period, $\frac{1}{2}T$, is given by

$$b=c\int_0^{\frac{1}{2}T}\beta dt$$

$$=\frac{c}{\omega}\int_0^{\pi}\frac{[A+B(1-\cos x)]dx}{\{1+[A+B(1-\cos x)]^2\}^{\frac{1}{2}}}. \tag{5}$$

This integral was integrated graphically to determine b for the first acceleration, letting A equal zero and B equal $\frac{1}{3}$. These values correspond to zero input velocity and a value of the electric field which was considered practical. The electrons are assumed to enter the cavity at the instant t equals zero; it is estimated that those electrons which enter the cavity by more than plus or minus five electrical degrees become so far out of phase with the accelerating field that they emerge from the cavity with energies differing by a factor of two or more from those which enter near the zero phase. The result of substituting t equal $\frac{1}{2}T$ in Eq. (4) yields a value of $\beta(1-\beta^2)^{-\frac{1}{2}}$ equal to $\frac{2}{3}$. This is the value of A used in computing the second value of b, for which B was again assumed to be $\frac{1}{3}$.

A model of the first two sections was constructed using the dimensions obtained above. Measurements showed that the field in the section corresponding to BC of Fig. 1 was half the value in the section corresponding to AB. The value of b for the second section was there-

[2] S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio* (John Wiley and Sons, New York, 1945), p. 389.

fore recomputed, using $\frac{1}{6}$ for the value of B of Eq. (5). Since this large change of B produced a change of b of only six percent, the new value of b was used in the final cavity. In the third section of the accelerator this value of B was assumed, and checked on the model. The model referred to here will be described later. This same model was used to determine the geometry of the bends for minimum reflection and to determine the positions of the holes such that the accelerating portions would be a wave-length apart.

In order that the cavity resonate as a single unit it is necessary that the bends introduce a minimum reflection. If these reflections are small their only effect will be to modify the relative magnitudes of the accelerating fields. Since these relative magnitudes were determined experimentally after the geometry of the bends had been fixed, no attempt was made to eliminate reflections entirely.

Since it was arbitrarily decided that the outside wall of the bend should be half of a right circular cylinder and the inside should have a radius of curvature of $\frac{1}{8}''$, there was left only one parameter for the design of the bends. This parameter was the spacing between the inner and outer curved walls. The model on which this dimension and the hole positions were determined is shown in Fig. 2. The wall spacings are made equal to the computed values of b for two adjacent sections of the cavity. Slots along C and F in the figure allowed measurements of the standing-wave pattern to be made with a hand probe. Power from a test oscillator was fed to the section at A. The plunger D was adjusted for a minimum standing-wave ratio along C when a matched load was placed at G. This procedure determined the spacing between inner and outer walls for the bend. For this adjustment a standing-wave ratio of approximately two to one was obtained.

Having fixed the dimensions of the bend it is necessary to fix the positions of the electron entrance and exit holes relative to the position of the bend. This was accomplished by replacing the matched load at G with a shorting plunger. The plunger position was changed until the minimum near F was in line with the minimum near C of Fig. 2. It was determined that these

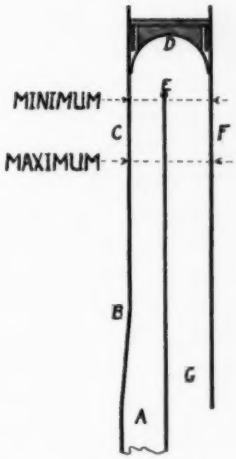


FIG. 2. Plan of test model, showing adjustable bend, D.

minima were a half-wave apart. The electron entrance and exit holes were then located each a quarter wave-length from the position of the minimum, thus insuring a separation of a whole wave-length. The ratio of the maxima obtained at this time was taken as the ratio of the fields in the final cavity. This measurement was used to determine the corrected value of parameter B in Eq. (5). For the second bend the dimensions of the second and third cavity sections were used in constructing a similar model, and the above measurements repeated. The computations of the widths of the accelerating sections and the measurements of the bend dimensions supply all data necessary to construct the cavity, with the exception of the inductive window.

The cavity was constructed of $\frac{1}{16}$-in. sheet copper which was silver-soldered along the

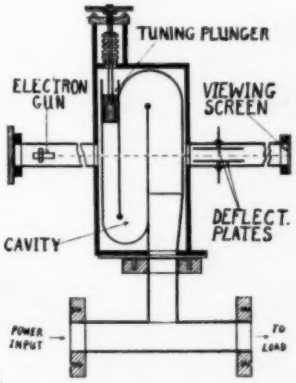


FIG. 3. Plan of cavity, showing vacuum jacket and auxiliary apparatus.

corners. A tapered section approximately a half-wave in length serves as the transition between the cavity and standard wave guide. The open end of the tapered section is covered by a $\frac{3}{16}$-in. polystyrene plate which serves as a vacuum seal. The inductive window was adjusted to give a standing-wave ratio of better than 1.1 to 1.0 with the polystyrene plate in place. The tuning plunger was adjusted inside the vacuum by a threaded rod soldered to a Sylphon bellows. The cavity was coupled to the magnetron by the series "Tee" arrangement employed by J. C. Slater.** The plan of the cavity, vacuum chamber, and associated apparatus is shown in Fig. 3.

OPERATION

With this cavity, 300 kv electrons were obtained. The energy of the electrons was measured by the electrostatic deflection method. The type 5CC-1, cathode-ray tube electron gun, which was used as a source, formed a collimated beam of electrons which were observed on a fluorescent screen on the opposite side of the cavity, as shown in Fig. 3. The high energy beam could be examined on the same screen and was found to form a spot of the same size as the low energy beam, indicating a high degree of bunching. In addition to the high energy beam a general background of illumination resulting from scattered electrons was observed. This may have been the result of the relatively high (10^{-5} mm Hg) pressure at which the system operated. The high energy beam was measured by comparing the voltages necessary to produce the same deflection of the high energy beam and the 2000-volt beam direct from the electron gun. From measurements of the average power dissipated by the cavity and a knowledge of the duty cycle (3.4-microsec. pulses at 60 pulses per sec.), the power averaged over a single pulse was calculated to be 250 kw when the desired accelerating field was obtained. The remainder of the 800 kw developed by the magnetron was dissipated in a water load.

The design procedure used above was simple and straightforward, and with obvious refinements could be applied to an accelerator of as many sections as the available power would allow. Although some other basic cavity arrangement may be more efficient, the system described has the advantage of simplicity, and is particularly applicable at low energy levels, where β is less than unity.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Professors L. B. Snoddy and J. W. Beams for their kind help and advice on this work, and to the members of the Laboratory shop, Mr. F. Linke and Mr. P. Sommer, for their construction of various parts of the apparatus.

** Private communication.

Temperature Coefficient of Electrical Resistivity for Crystalline Selenium Containing Various Percentages of Bromine

WAYNE E. BLACKBURN
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania
(Received July 15, 1947)

The temperature coefficient of electrical resistance of brominated selenium over the temperature range −30°C to 100°C was found to be negative and variable. The electrical resistivities of the selenium samples varied by factors of three to twelve for the temperature range mentioned, depending upon the bromine content. A reproducible hysteresis loop was obtained on the plot of resistivity *versus* temperature. A minimum in the plot of electrical resistivity *versus* bromine content was found with 0.007 percent bromine in selenium. Activation energies for freeing electrons obtained from the slopes of conductivity *versus* temperature plots indicate a dependence upon bromine content in the selenium.

I. INTRODUCTION

SELENIUM is a non-metallic element which exists in several allotropic forms. It was discovered in 1817 by J. J. Berzelius and J. G. Gahn. The element is usually obtained in a red amorphous form from the flue dust of pyrites burners. This material melts at 217°C and, if kept somewhat below that temperature, will become crystalline. In the latter state it has a gray metallic appearance and is electrically conductive. Its electrical resistance depends not only upon temperature but also upon the amount of light to which it is exposed, which fact furnishes the basis for its use in photoelectric cells. The element resembles sulfur and has valencies of two, four, and six.

This investigation covers the temperature range −30°C to 100°C, in which the electrical conductivity of selenium is determined as a function of temperature and bromine content.

II. THEORY

The present theory of electron conduction[1] assumes quantized energy levels for the electrons of a single isolated atom. Pauli's exclusion principle states that not more than two electrons can occupy the same energy level of an atom, and, if there are two, they must have their spins in opposite directions. Since this principle applies to crystals as well as to single atoms, quantum theory states that there exist in crystals energy states occurring in bands of very closely spaced levels, the bands being separated by regions called forbidden zones. If the number of valence electrons is such as just to fill an energy band, the next allowable energy levels are in a higher band at a height such that an applied field strength of the order of 10^6 v/cm would be required to transfer electrons to this band. This value of field strength is known as the breakdown potential. Such a crystal behaves as an insulator. If an energy band is not completely full of electrons, an electric field can give electrons sufficient energy to pass up into higher conduction levels, thereby producing an electric current. Such a crystal is a metal. Anomalous behavior is observed if a band is nearly full of electrons. Electrical conduction is then described in part in terms of motion of vacant spaces or "holes" which act like positively charged electrons. Measurements of the Hall effect and thermoelectric effect will determine in what way a given crystal is carrying current, that is, by positive "holes" or by electrons.

There are two kinds of semiconductors,[2] the

TABLE I.
Analysis of double-distilled selenium.*

Selenium	99.9676 percent
Tellurium	.026
Iron	.0027
Silica	.002
Lead	.0001
Copper	.00012
Sulfur	.0014

* Material and analysis supplied by Canadian Copper Refineries Limited, Montreal East, Quebec.

[1] A. H. Wilson, *Semiconductors and Metals* (Cambridge University Press, Teddington, England, 1939).

[2] N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Oxford University Press, London, 1940), p. 152.

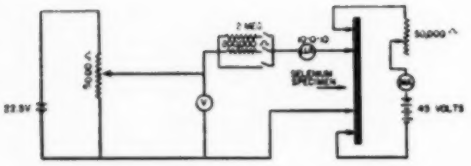


FIG. 1. Resistivity measuring circuit.

intrinsic and the extrinsic. An intrinsic semiconductor is one which conducts in the pure state; electrons are evaporated by thermal energy into the conduction levels from the normally full band. A semiconductor usually owes its conductivity to the presence of impurities, and its actual value is very sensitive to the amount of impurities present. An intrinsic conductivity which is proportional to the number of electrons that possess sufficient thermal energy to pass from the lower normally full band into the closest upper normally empty band, is given by an exponential function

$$\sigma = \sigma_0 \exp\left[-\frac{\epsilon}{kT}\right],$$

where σ represents the conductivity, σ_0 is a constant, ϵ the energy difference between the bands, known as the activation energy, k the Boltzmann atomic constant, and T the absolute temperature. The introduction of impurities into such a semiconductor may introduce energy levels in the normally forbidden region between the two bands mentioned and thus changes the value of ϵ. The new ϵ is now equal to the energy difference between the energy levels of the impurity and the closest band of energy levels in the semiconductor. In this way the energy step may be made small enough to permit electrons to be evaporated up into higher energy level because of thermal excitation. If the energy levels of the impurity are just above the top filled band, these impurity levels accept electrons from this full band of the semiconductor. Holes are thereby created in this normally full band, their number depending, among other things, upon the value of ϵ. Current is then carried in large part by the so-called "positive holes" produced in the full band. Such a semiconductor is known as a "defect conductor." Selenium is known to be one of this type. If the energy levels of the impurity are just below an empty conduction band these impurity levels donate electrons to the empty conduction band, making electrical currents possible. Such a semiconductor is known as an "excess conductor."

Considering a practical case,[3] the number of conduction electrons in an extrinsic conductor increases with increasing temperatures because bound electrons are freed by thermal agitation. The result is increased conductivities at higher temperatures. This increase in number of electrons is considered to follow the exponential function previously mentioned. The ϵ, previously described, is now considered the energy with which the electrons are bound to internally adsorbed atoms, to colloidal impurities or to imperfections of the lattice. In other words, it is the energy which is necessary to transfer the electron from the place where it is bound into a higher energy level.

Theoretical treatment[4] indicates that when $\log\sigma$ is plotted against $1/T$ a straight line should result whose slope is proportional to the activation energy, ϵ. This assumes $n \ll Ne$, where n is the number of free electrons that have been evaporated into the conduction band and Ne the maximum number of electrons which can be excited from the active impurity levels.

The purpose of this study is to determine how the electrical conductivity of selenium changes with temperature and bromine content. Also by applying the theory presented the activation energies can be computed for selenium containing known amounts of impurity.

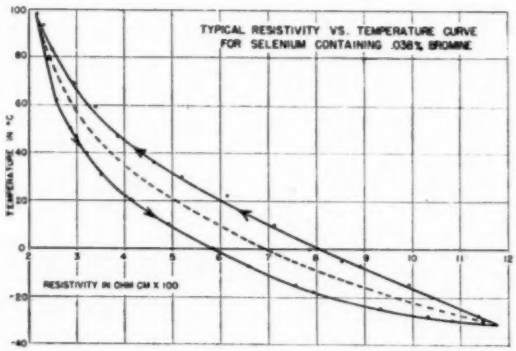

TYPICAL RESISTIVITY VS. TEMPERATURE CURVE
FOR SELENIUM CONTAINING .038% BROMINE
TEMPERATURE IN °C
RESISTIVITY IN OHM CM X 100

FIG. 2.

[3] J. H. DeBoer, *Electron Emission and Adsorption Phenomena* (The Macmillan Company, New York, 1935), pp. 311–317.

[4] N. F. Mott and R. W. Gurney, see reference 3, p. 160.

III. APPARATUS AND METHOD

A. Preparations of Selenium Specimens

Double-distilled, high purity selenium (see Table I) was melted and heated to a temperature of 400°C. It was then poured onto a flat Pyrex glass plate which was at 240°C. A framework was previously placed on the glass plate limiting the flow of selenium to a definite shape. The selenium and glass plate were allowed to cool to room temperature requiring about 25 minutes. The mold was then placed in a 207°C oven and heat treated for 90 minutes. In the heat-treating oven the insulating amorphous selenium became crystalline and conductive; it was removed and permitted to cool to room temperature. The selenium was then in a dense uniform sheet about 0.080 inch thick. Care was taken to have the Pyrex glass plate chemically clean and the heat treatments follow the prescribed time, otherwise a deep wrinkled surface results on the selenium-glass face. Also spongy crystalline selenium resulted if the annealing was carried out too slowly. The uniform sheets of selenium were cut into specimens 3.25 inches long, 0.50 inch wide, and about 0.080 inch thick.

The various percentages of bromine in the prepared specimens were obtained by making a 1.0 percent concentrate of selenium tetrabromide in high purity selenium. This concentrate was then used to make the desired bromine percentages in subsequent samples by dilution. Several specimens from each bromine concentration were chemically analyzed to determine the percentage of bromine in the finished specimen.

B. Chemical Analysis of Treated Selenium

The crystalline selenium samples were dissolved in dilute nitric acid. This causes hydrolysis

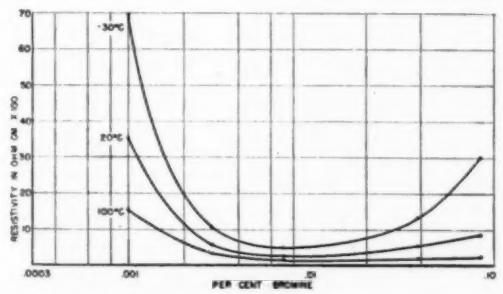


FIG. 3. Resistivity of selenium *vs.* percent bromine.

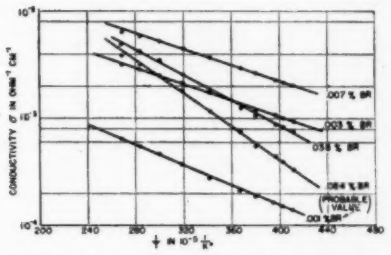


FIG. 4. Conductivity *vs.* reciprocal of absolute temperature for selenium.

of selenium halides to hydrogen halides. During the solution procedure some halogen may be lost by volatilization of either selenium halides or hydrogen halides. This was prevented by using an all-glass distillation apparatus with a reflux condenser, to the exit of which were attached, in series, two vessels containing a solution which absorbed such halides.

After reaction was complete, the absorbent liquids were added to the main body of solution, a slight excess of silver nitrate was added, and the test solution boiled. This causes any precipitated silver selenite to redissolve. The solution was now allowed to stand until coagulation of silver halide was complete. The precipitate was then filtered, dried, and weighed.

This method is quite accurate for determination as low as 0.01 percent bromine in selenium, using 25-gram samples.

C. Electrical Conductivity Measurements

The electrical conductivity as a function of temperature was measured on all specimens by the following method: A heavy continuous coat of graphite was put on the ends of the specimen by rubbing with a pencil, this serving as good electrical contact for current supply terminals. The specimen was then placed in a holder having four nickel-plated bronze electrodes. Two terminal current supply leads and two center voltage probes were provided. A thermocouple was placed in the plane of the base surface of the holder adjacent to the specimen. The holder was submerged in a bath of silicone oil, and the current supply leads were connected to a d.c. source while the center voltage probes were connected to a *K*-type potentiometer circuit as shown in the accompanying Fig. 1. The container with oil and the specimen holder were placed in a tempera-

TABLE II.
Activation energies of brominated selenium obtained from slopes of curves, Fig. 4.

Percentage bromine* (±.0005)	Activation energy (electron volts) (±.002)
.0070	0.074
.0030	0.077
.0380	0.121
.0840	0.155
.0010 (probable content)	0.091

* Analysis of selenium samples by Dr. H. L. Wunderly, Analytical Section, Westinghouse Research Laboratories; method of analysis developed by Dr. R. A. Bournique.

ture-controlled box. A direct current of about one milliampere was passed through each specimen, and the temperature was varied from −30° to 100°C.

IV. DISCUSSION OF RESULTS

A typical curve obtained by plotting resistivity *versus* temperature is shown in Fig. 2. The curve shifts to higher resistance after having been cooled to −30°C the first time. The extent of this shift is proportional to the initial resistance of the specimen. The cycle, −30°C to 100°C, is repeatable after the first thermal shock. All specimens showed a hysteresis loop in traversing a complete cycle. The width of the loop is a function of the cold terminal point; that is, if −30°C were exceeded, the hysteresis loop would be wider, but a common trace always occurs at about 110°C. Indications are that this is an inherent property of brominated selenium. The values selected for the electrical resistivity of a given specimen were taken from the dotted curve shown bisecting the loop and used for the purpose of relating the resistivity as a function of bromine content. Data obtained were based on the mean of several specimens of each composition. Figure 3 shows resistivity *versus* percentage bromine for different temperatures. A minimum resistance occurs for a bromine content of about 0.007 percent bromine. The explanation might possibly be that the bromine ion density became such that recombination began, forming inactive bromine molecules.

With reference to theory presented, the activation energy, E, was determined with respect to bromine content. Figure 4 shows $\log\sigma$ *versus* $1/T$ for different bromine percentages. From the slopes of the curves, and by assuming the exponential law to hold, the activation energies in electron volts were computed (see Table II). For example,

$$\log\sigma = \log\sigma_0 - \frac{E}{k}\left(\frac{1}{T}\right),$$

and for the case of 0.007 percent bromine in selenium,

$$\frac{E}{k} = \frac{\ln(0.060/0.022)}{(400-283)\times 10^{-5} 1/K^\circ} = 8.55\times 10^2 K^\circ,$$

$$E = 8.55\times 10^2 K^\circ \times 8.69\times 10^{-5} \text{ev}/K^\circ = 0.074\text{ev}.$$

ACKNOWLEDGMENT

The author wishes to express his appreciation to Dr. E. D. Wilson and Mr. G. W. Penney of the Westinghouse Research Laboratories for suggestions and counsel, and the Westinghouse Electric Corporation for financial assistance.

The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers

W. C. ELMORE*
University of California, Los Alamos Scientific Laboratory, Santa Fe, New Mexico
(Received April 11, 1947)

When the transient response of a linear network to an applied unit step function consists of a monotonic rise to a final constant value, it is found possible to define *delay time* and *rise time* in such a way that these quantities can be computed very simply from the Laplace system function of the network. The usefulness of the new definitions is illustrated by applications to low pass, multi-stage wideband amplifiers for which a number of general theorems are proved. In addition, an investigation of a certain class of two-terminal interstage networks is made in an endeavor to find the network giving the highest possible gain—rise time quotient consistent with a monotonic transient response to a step function.

1. INTRODUCTION

THE transient behavior of any linear system (or network) is contained implicitly in the *system function* $F(s)$ which expresses directly the steady-state (sinusoidal) response of the system. The variable in the system function, $s=\sigma+j\omega$, is the complex angular frequency; ω is the ordinary (real) angular frequency, and σ is a real variable introduced for the purpose of facilitating the transient analysis of the system.[1] In the present paper we shall be concerned primarily with the class of linear systems in which the transient response to a unit step function (the so-called *indicial admittance*) consists of a *monotonic* rise to a final constant value. For simplicity in presentation only the transient response of a low pass, wideband amplifier will be considered. Many of the results obtained, however, apply equally well to other electrical systems, as well as to mechanical, acoustical, thermal, and to mixed systems, provided only that they are linear and have a monotonic transient response to a unit step function.

The most important system function of an amplifier is the complex gain, $G(s)$, connecting input and output voltages of the form $E\epsilon^{st}$. In the case of a low pass amplifier, $G(s)$ can always be separated into two factors, $G_1(s)$, which governs the response at low frequencies, and $G_2(s)$, which governs the response at high frequencies. In an unfedback amplifier, $G_1(s)$ owes its origin to various RC networks which couple the plate of one tube to the grid of the next, and which furnish bias voltages to various points in the amplifier. The system function $G_2(s)$ owes its origin primarily to parasitic interstage capacitances which shunt the signal-carrying leads. Since we shall be interested in the problem of obtaining the greatest possible *gain—rise time quotient* for an amplifier,[2] $G_2(s)$ may reflect the presence of compensating inductances, of feedback, or of any other circuit arrangements used to shorten the rise time or to improve the transient properties of the amplifier. The portion $G_2(s)$ of the system function may be considered as that of an equivalent amplifier idealized to have perfect low frequency response.

For convenience in analysis, we shall use the *normalized* system function $g_2(s)=G_2(s)/G_2(0)$, where $G_2(0)$ is the gain of the idealized amplifier at zero frequency. Normalization evidently makes the final value of the response to a unit step function (given by the final-value theorem of the Laplace transformation) also unity.

It is not difficult to show that the normalized system function $g_2(s)$ of a stable amplifier containing a finite number of lumped circuit elements takes the form

$$g_2(s)=\frac{1+a_1s+a_2s^2+\cdots+a_ns^n}{1+b_1s+b_2s^2+\cdots+b_ms^m}, \qquad (1)$$

* Now at Swarthmore College, Swarthmore, Pennsylvania.

[1] The notation and terminology adopted here is that found in M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, (John Wiley and Sons, Inc,. New York, 1942), Vol. 1.

[2] The gain—rise time quotient is analogous to the more familiar gain-band width product, but appears to be a more useful measure of amplifier performance in the case of amplifiers designed to amplify fast transients. The definition of rise time is considered in Section 2.

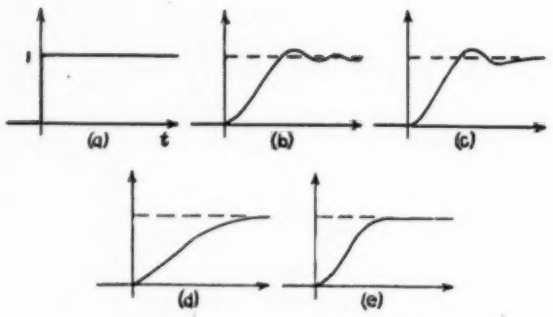


FIG. 1. Some typical transient response curves.

where the coefficients a_i and b_i are real, $m>n$ and the poles of $g_2(s)$ all lie in the left half of the complex s-plane. The normalized transient response of the amplifier to the unit step function $u(t)$ can be computed by means of the inverse Laplace transformation

$$e(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{1}{s}g_2(s)\epsilon^{st}ds, \quad c>0. \qquad (2)$$

Transient response curves computed from Eq. (2) for various amplifiers have a variety of shapes, some common forms of which are illustrated in Fig. 1. The input signal, $u(t)$, is shown in (a). The transient response shown in (b) consists of a delayed rise, followed by a train of damped oscillations. The response shown in (c) is similar to that in (b) except that only a finite number of oscillations occur, preceding a gradual approach of the curve to the final value unity. In (d) and (e) are illustrated monotonic transient response curves having different amounts of damping. The response in (e) is supposed to be that of an amplifier having certain adjustable circuit parameters which have been chosen to achieve the shortest possible monotonic rise for a given amplifier gain.

Any circuit elements introduced in an amplifier for the purpose of controlling the shape of the transient response curve may be termed *compensating elements*. In the present instance they afford high frequency compensation to the response of the amplifier. When the fastest possible monotonic rise has been obtained with the particular type of compensation used, the amplifier is said to be *critically compensated*. If the transient response is monotonic but the rise is slower than can be obtained by suitably adjusting the compensating elements, the amplifier will be referred to as *under-compensated*. Finally, if the transient response is not monotonic, usually as the result of increasing the compensating elements beyond the point giving critical compensation, the amplifier will be referred to as *over-compensated*.[3]

It is evident that the various types of transient curves illustrated in Fig. 1 possess certain common features, in particular, a delay which occurs before the response is well under way, and a finite time of rise. For many purposes each curve can be sufficiently well characterized by its *delay time* and *rise time*, which can be defined in several different, but approximately equivalent ways. One of the purposes of the present paper is to propose useful definitions for these quantities, with a view to facilitating their computation from the system function $g_2(s)$. The new definitions, unfortunately, are of such a nature that they apply only to systems which are not over-compensated. Their utility for all systems having a monotonic transient response, however, appears to be great enough to outweigh this defect. It is possible that an equally useful method for treating the over-compensated case can be discovered.

2. THE DEFINITION OF DELAY TIME AND OF RISE TIME

A number of definitions of delay time and of rise time appear to be in practical use. Two of these will be illustrated by reference to Fig. 2, which shows the transient response $e(t)$ to the unit step function, and its derivative, $e'(t)$, of an under-compensated amplifier.[4]

The delay time, T_D, is usually defined as the time required for the response to reach half its final value, as illustrated in Fig. 2a. The rise time, T_R, is sometimes defined as the reciprocal of the slope of the tangent drawn to the response curve at its half-value point, again as illustrated in Fig. 2a. A somewhat more practical definition

[3] For many applications it is important to avoid over-compensation in an amplifier. This is particularly true for *pulse amplifiers* used in nuclear physics (to amplify pulses obtained from an electrical detector of radiation) and for wideband amplifiers used in studying fast electrical transients (such as amplifiers for cathode-ray oscillographs). Video amplifiers used in television applications are not as critical in this respect since small oscillations resulting from over-compensation do not impair the quality of television pictures.

[4] The curve $e'(t)$ may be considered to be the response of the amplifier to a *unit impulse* applied at time $t=0$.

results if T_R is taken to be the time required for the response to increase from 10 to 90 percent of its final value.[5] Although these definitions are useful in the laboratory, they are extremely awkward for making computations, or for entering upon a theoretical investigation of the relative merits of various methods of compensating an amplifier to reduce its rise time. The difficulty, of course, lies in the necessity for computing the transient response curve for each case under consideration, a formidable undertaking. It is practically impossible to obtain values of T_D and T_R, as defined, by any simple method of analysis.

Let us now consider alternative definitions for delay time and rise time. Evidently the delay time should be measured from $t=0$ to some time at which the transient rise is about one-half over. It is reasonable, therefore, to measure T_D to the centroid of area of the curve $e'(t)$, that is,

$$T_D=\int_0^\infty te'(t)dt. \tag{3}$$

The formula for the centroid takes this simple form since

$$\int_0^\infty e'(t)dt=1.$$

This definition of delay time is illustrated in Fig. 2b, and it is seen to give a result which differs but little from that obtained from the customary definition. The two values of delay time depart most markedly in the case of a very asymmetrical response curve. It is easy to convince oneself that the new definition becomes meaningless if the curve $e'(t)$ possesses a negative portion, i.e., $e(t)$ is not monotonic. It will be shown presently that it is a simple matter to obtain a value of the integral in Eq. (3) directly from the system function $g_2(s)$.

The rise time T_R should express in a prescribed manner the time required for the transient rise to occur. Now the shorter the rise time, the narrower (and higher) the curve of $e'(t)$. It is reasonable, therefore, to define T_R as proportional to the radius of gyration of the area under the curve, that is,

$$T_R^2=\text{Const.}\int_0^\infty (t-T_D)^2e'(t)dt. \tag{4}$$

In other words, the rise time is proportional to the standard deviation of the response $e'(t)$. The constant of proportionality is chosen to be 2π for the following reason: it is possible to show that the curve $e'(t)$ for any n-stage amplifier[6] approaches more and more closely the form of a Gauss error curve with increasing n.[7] To make the new definition of rise time agree with the definition based on the slope of the transient response curve (Fig. 2a), the value of T_R should therefore be

$$T_R=\frac{1}{e'(t)\,|_{\max}}=(2\pi)^{\frac{1}{2}}\,[\text{radius of gyration of } e'(t)],$$

which expresses the relation between the height and the radius of gyration of a Gauss error curve of unit area, here denoted temporarily by $e'(t)$. Equation (4) can now be written

$$T_R=\left\{2\pi\left[\int_0^\infty t^2e'(t)dt-T_D^2\right]\right\}^{\frac{1}{2}}, \tag{5}$$

where the integral has been expressed in terms of moments about the time origin. It is found in most instances that rise times computed from Eq. (5) differ by less than ten percent from the

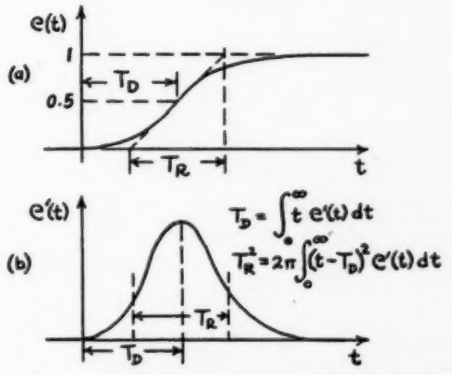


FIG. 2. Curves illustrating the definitions of delay time and of rise time.

[5] H. E. Kallman and R. E. Spencer, Proc. I.R.E. **33**, 169–195 (1945).

[6] The individual stages in the amplifier must each have a monotonic transient response to the unit step function.

[7] This result appears to have been first noticed by Henry Wallman, and will be discussed in Chap. 7, Vol. 18 of the *Radiation Laboratory Series* (McGraw-Hill Book Company, Inc., New York, in press).

rise times defined earlier, which can continue to be used for most laboratory work.

The great usefulness of the new definitions of delay time and rise time will now be demonstrated. The system function $g_2(s)$ and the transient response $e'(t)$ are related by the direct Laplace transformation

$$g_2(s)=\int_0^\infty e'(t)\epsilon^{-st}dt, \tag{6}$$

where σ, the real part of s, must be greater than σ_1, the real part of the pole s_1 of $g_2(s)$ lying farthest to the right in the s-plane. In the case of a stable amplifier, σ_1 is *negative*, in fact, for a monotonic transient response, the poles of $g_2(s)$ all lie on the negative real axis of the s-plane. Let us now expand the Laplace integral (6) in a power series in s, which will be a valid expansion of $g_2(s)$ for values of s lying within the circle of convergence $|s|=|s_1|$. We do this by first expanding ϵ^{-st} in a power series in st and then integrating term by term, obtaining the development

$$g_2(s)=1-s\int_0^\infty te'(t)dt+\frac{s^2}{2!}\int_0^\infty t^2e'(t)dt-\cdots. \tag{7}$$

It follows that if a given system function is expanded in ascending powers of s, it is a simple matter to obtain by inspection the first and second moments of $e'(t)$ about the time origin,[8]

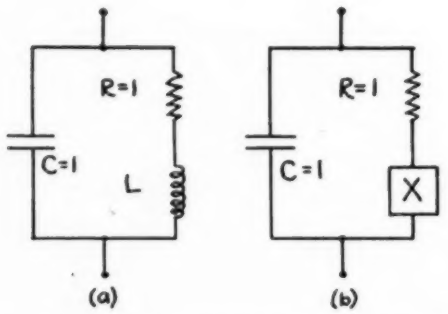


FIG. 3. Shunt-compensated interstage networks.

and therefore to obtain values of T_D and T_R defined by Eqs. (3) and (5), respectively. Part of the virtue of the proposed definitions lies in the ready way in which delay times and rise times can be computed. Other advantages of the definitions will be made use of in Section 3.

It is useful to obtain expressions for T_D and T_R for a system function of the form given by Eq. (1). By expanding Eq. (1) in ascending powers of s, it is found that

$$T_D=b_1-a_1, \tag{8}$$

and that

$$\frac{T_R^2}{2\pi}=b_1^2-a_1^2+2(a_2-b_2). \tag{9}$$

Before considering other matters, let us compute values of T_D and T_R for a single-stage amplifier having a two-terminal plate load impedance of the type shown in Fig. 3a.[9] Such an amplifier stage is said to be *shunt compensated*. The system function of the single stage is identical to the driving point impedance of the plate load, since it can be assumed that the amplifying tube is equivalent to a constant current generator. Hence we have that

$$g_2(s)=\frac{1+Ls}{1+s+Ls^2}. \tag{10}$$

In order that no transient oscillation of the type shown in Fig. 1b shall exist, the poles of $g_2(s)$ must lie on the negative real axis of the s-plane. This requires that in Eq. (10) $L\leq 1/4$. The values of T_D and T_R (computed using Eqs. (8) and (9)) are

$$\left.\begin{aligned} T_D&=1-L,\\ T_R&=(2\pi(1-2L-L^2))^{\frac{1}{2}}.\end{aligned}\right\} \tag{11}$$

When $L=0$, corresponding to a simple resistance-coupled amplifier stage, $T_D=1$, and $T_R=(2\pi)^{\frac{1}{2}}=2.51$. When $L=1/4$, corresponding to critical shunt compensation, $T_D=3/4$, and $T_R=(2\pi)^{\frac{1}{2}}(7/16)^{\frac{1}{2}}=1.66$. To express the improve-

[8] It has been called to the attention of the writer that this method of computing moments is closely related to methods used in mathematical statistics, the method of the moment generating function, and the method of the characteristic function. See, for instance, S. S. Wilks, *Mathematical Statistics* (Princeton University Press, Princeton, New Jersey, 1943).

The mathematical steps leading to Eq. (7) can be made more rigorous by noticing that the series for the integrand converges *uniformly*, thus permitting term-wise integration between zero and a finite upper limit T. Since the integrated series, considered as a function of T, is also uniformly convergent when s is restricted to the region near the origin, the series must converge, as T approaches infinity, to the Laplace integral (6) from which it is derived.

[9] By setting $C=1$, $R=1$, and expressing L in units of R^2C, values of T_D and T_R are obtained in units of RC. This device enables the system function to be written immediately in a simple, normalized form.

ment realized by compensating the stage it is convenient to define the *rise-time figure-of-merit* S, which is the ratio of the rise time of an R-coupled stage to the rise time of the same stage (i.e., a stage with the same value of R and C) compensated to reduce its rise time. Evidently $S=(16/7)^{\frac{1}{2}}=1.51$ for a critically shunt-compensated stage. In Section 4 an attempt will be made to discover an interstage network which gives the smallest value of T_R with a given interstage parasitic capacitance and load resistance, that is, the largest value of S. The problem is somewhat analogous to that of discovering the network which leads to the maximum band width (without regard to good transient response).[10]

3. SOME THEOREMS REGARDING MULTI-STAGE AMPLIFIERS

We have just seen how the delay time and rise time of a single amplifier stage can be computed. Let us now consider how the delay time and rise time of an unfedback multi-stage amplifier depends on the properties of individual stages in it.

If the amplifier contains n stages in tandem, the system function of the entire amplifier is the product of the system functions of the individual stages.[11] Let the system function of the ith stage be $g_{2i}(s)$, and let the corresponding values of delay time and rise time be T_{Di} and T_{Ri}, respectively. The function $g_{2i}(s)$ can be expanded in the series

$$g_{2i}(s)=1-sT_{Di}+\frac{s^2}{2}\left(\frac{T_{Ri}^2}{2\pi}+T_{Di}^2\right)-\cdots, \quad (12)$$

which is obtained directly from Eqs. (3), (5) and (7). The system function of the entire amplifier therefore becomes

$$g_2(s)=\prod_1^n g_{2i}(s)$$
$$=1-s\sum T_{Di}+\frac{s^2}{2}\left[\sum\frac{T_{Ri}^2}{2\pi}+\sum T_{Di}^2+2\sum_{i>j}T_{Di}T_{Dj}\right]-\cdots.$$

By again using Eqs. (3), (5) and (7), the delay time and the rise time of the entire amplifier are found to be

$$T_D=\sum_1^n T_{Di}, \quad (13)$$

and

$$T_R=\left(\sum_1^n T_{Ri}^2\right)^{\frac{1}{2}}. \quad (14)$$

The result expressed by Eq. (13) is intuitively obvious, since it is to be expected that the total delay is the sum of the delays of the individual stages. The manner of combining rise times indicated by Eq. (14) is not as evident, although the fact that this simple mode of combination is the correct one has been proposed by several of the author's colleagues prior to the present proof of the theorem.

Another theorem of practical importance concerns the manner in which the gain of an n-stage amplifier should be distributed among the individual stages in order to achieve the shortest possible over-all rise time for a given over-all gain. Now the rise time of any stage in the amplifier varies directly with the gain of the stage, since both quantities are proportional to the value of the plate load resistor. It is desired, therefore, to minimize the expression (14) subject to the condition that

$$\prod_1^n T_{Ri}=\text{Constant}. \quad (15)$$

It is easy to prove from Eqs. (14) and (15) that the over-all rise time is a minimum when the rise times of all stages are made the same.[12] If T_{R1} is the rise time of each stage, the rise time of an n-stage amplifier becomes

$$T_R=T_{R1}n^{\frac{1}{2}}. \quad (16)$$

Let us now consider certain matters regarding the design of an amplifier consisting of n identical stages. We shall treat the simple case where the interstage couplings are of the general type

[10] See, for instance, H. W. Bode, *Network Analysis and Feedback Amplifier Design*, (D. Van Nostrand Company, Inc., New York, 1945), pp. 408, *et seq.*

[11] This statement is true provided that no coupling between stages exists except through the electron stream in the constituent amplifying tubes. This situation can be realized in practice if the tubes in the amplifier are pentodes.

[12] For instance, by using Lagrange's method of undetermined multipliers, the differential of Eq. (14) must be zero, $\Sigma\, T_{Ri}dT_{Ri}=0$, subject to the condition $(\Pi\, T_{Ri})\,\Sigma\, dT_{Ri}/T_{Ri}=0$ which is the differential of Eq. (15). After multiplying the latter equation by the undetermined multiplier α and adding it to the former, each coefficient of dT_{Ri} must be identically zero, giving $T_{R1}=T_{R2}=\cdots$. A proof that this condition leads to a *minimum* rise time is scarcely needed.

illustrated in Fig. 3, i.e., a parasitic capacitance C, and a resistance R in series with some sort of compensating reactance whose impedance becomes zero at zero frequency. At frequencies where $1/\omega C \ll R$, the gain of each stage is

$$G_1 = g_m R, \tag{17}$$

where g_m is the transconductance of the amplifying tube.[13] The rise time of each stage can be written in the form

$$T_1 = \frac{1}{S}(2\pi)^{\frac{1}{2}} RC, \tag{18}$$

where S is the rise-time figure-of-merit of the stage. By definition $S=1$ for a simple R-coupled stage and we have already shown, for example, that $S=(16/7)^{\frac{1}{2}}$ for a critically shunt-compensated stage. Eliminating the resistance R between Eqs. (17) and (18), we have that

$$G_1 = S\left(\frac{g_m}{(2\pi)^{\frac{1}{2}} C}\right) T_1. \tag{19}$$

The quantity $g_m/(2\pi)^{\frac{1}{2}}C$ expresses the figure-of-merit of the amplifier tubes, and may be conveniently stated as so much gain per microsecond rise time.

If T is the rise time of the n-stage amplifier, then according to Eq. (16) the rise time of each stage must be $T_1 = T/n^{\frac{1}{2}}$, requiring a gain for each stage

$$G_1 = S\left(\frac{g_m}{(2\pi)^{\frac{1}{2}} C}\right) \frac{T}{n^{\frac{1}{2}}}. \tag{20}$$

Equation (20) can be written as a pair of equations,

$$\left.\begin{aligned} G_1 &= \frac{S}{n^{\frac{1}{2}}} G_0, \\ \text{where} \\ G_0 &= \left(\frac{g_m}{(2\pi)^{\frac{1}{2}} C}\right) T. \end{aligned}\right\} \tag{20a}$$

The quantity G_0 is the gain of a single R-coupled stage of rise time T, whereas G_1 is the gain that each stage of the n-stage amplifier must have in order that the entire amplifier shall have the rise time T.[14]

The total gain of the amplifier is

$$G_t = G_1^n. \tag{21}$$

Let us now investigate what gain per stage will result in the shortest rise time for a given total amplifier gain. From Eqs. (20) and (21), we have that

$$T = \frac{1}{S}\left(\frac{(2\pi)^{\frac{1}{2}} C}{g_m}\right) n^{\frac{1}{2}} G_t^{1/n}. \tag{22}$$

On minimizing T with respect to n while keeping G_t constant, it is found that $n = 2 \ln G_t$, or that

$$G_1 = \epsilon^{\frac{1}{2}} = (2.72\cdots)^{\frac{1}{2}} = 1.65\cdots. \tag{23}$$

This result is independent of the degree of compensation used, provided, of course, that critical compensation is not exceeded. The minimum rise time which can be obtained for a total gain G_t is found from Eqs. (20), (21), and (23) to be

$$T_{\min} = \frac{1}{S}\left(\frac{(2\pi)^{\frac{1}{2}} C}{g_m}\right) (2\epsilon \ln G_t)^{\frac{1}{2}}, \tag{24}$$

requiring a total of $n = 2 \ln G_t$ stages.

4. SOME CRITICALLY COMPENSATED INTERSTAGE NETWORKS

There are two matters of considerable interest concerning interstage networks of a critically compensated wideband amplifier. The first is primarily of theoretical importance and concerns the maximum value that can be obtained for the quantity S (the ratio of the rise time obtained with a simple RC network to that obtained with a compensated network). The second matter is of practical importance and concerns the design of

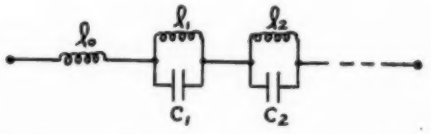


FIG. 4. A general driving point reactance having zero reactance at zero frequency. The inductance l_0 will vanish if the reactance must be zero at infinite frequency.

[13] It is assumed that $R \ll r_p$, the plate resistance of the tube.

[14] The pair of equations (20a) can be made the basis of a convenient nomograph to aid in the design of an amplifier of assigned rise time and total gain.

networks whose performance approaches as nearly as possible the theoretical limit.

Interstage coupling networks of two types must be distinguished, two-terminal and four-terminal.[15] This distinction is necessary since it is possible to separate the parasitic interstage capacitance into two portions, the output capacitance of one stage, and the input capacitance of the following stage. If a critically compensated four-terminal network is based on the two capacitances, as separate entities, it would seem likely that a shorter rise time can be achieved than for the two-terminal case. Only two-terminal networks of a simple type will be discussed in the present paper, mainly because a treatment of other cases is beset with algebraic difficulties.

Let us then consider the generalized, shunt-compensated interstage network illustrated in Fig. 3b, where the pure reactance X has a value zero at zero frequency, but is otherwise unrestricted in form. According to Foster's reactance theorem,[16] a possible formula for any reactance of this type can be written

$$X(s)=k\frac{s(s^2-s_2^2)(s^2-s_4^2)\cdots(s^2-s_m^2)}{(s^2-s_1^2)(s^2-s_3^2)\cdots(s^2-s_{m-1}^2)}, \quad (25)$$

where the s_i^2 $(i=1\cdots m)$ are negative real numbers, k is a positive constant, and m is an even integer. The general reactance can be realized physically by a variety of equivalent networks made up of inductances and capacitances.[16] It is convenient here to adopt the form of network shown in Fig. 4 to represent the general reactance $X(s)$. If the inductance l_0 vanishes, i.e., the general reactance becomes zero at infinite frequency, it is necessary to omit the factor $(s^2-s_m^2)$ from the right-hand member of Eq. (25).

TABLE I. Some critically compensated networks.

p	Circuit constants	S	Response $e'(t)$ to unit impulse
1	$X=0$	1.000	e^{-t}
2	$l_0=1/4$	1.512	$e^{-2t}(1+2t)$
3	$l_1=8/27;\ c_1=1/8$	1.769	$e^{-3t}(1+2t+6t^2)$
4	$l_0=1/4;$ $l_1=1/16;\ c_1=1$	1.899	$e^{-4t}(1+4t+(64/3)t^3)$
5	$l_1=(4/125)(5+(5)^{\frac{1}{2}})$ $c_1=(1/16)(3+(5)^{\frac{1}{2}})$ $l_2=(4/125)(5-(5)^{\frac{1}{2}})$ $c_2=(1/16)(3-(5)^{\frac{1}{2}})$	1.970	$e^{-5t}(1+4t+20t^2-(100/3)t^3+(250/3)t^4)$
∞	$l_k=2/\pi^2k^2$ $c_k=1/2$ $(k=1, 2, 3, \cdots)$	2.121	$\begin{cases}1-t/2, & 0\le t\le 2\\ 0 & 2\le t\end{cases}$

By writing Eq. (25) in the form

$$X(s)=\frac{d_1s+d_3s^3+\cdots+d_{m+1}s^{m+1}}{1+c_2s^2+c_4s^4+\cdots+c_ms^m},$$

where the new, real, positive constants, c_i and d_i, are uniquely related to the constants appearing in Eq. (25), and to the circuit constants defined in Fig. 4, we find that the driving point impedance becomes

$$Z(s)=\frac{1+d_1s+c_2s^2+d_3s^3+\cdots+c_{n-2}s^{n-2}+d_{n-1}s^{n-1}}{1+s+(d_1+c_2)s^2+c_2s^3+\cdots+c_{n-2}s^{n-1}+d_{n-1}s^n}, \quad (26)$$

where $n\equiv m+2$. It should be noticed that when $l_0=0$, the coefficient $d_{m+1}=d_{n-1}$ vanishes. Equation (26), of course, has the form of Eq. (1).

To realize a monotonic transient response to the unit step function, it appears necessary to require that the poles of $Z(s)$ all lie on the negative real axis of the s-plane. (Otherwise the transient response will contain oscillatory terms.) We shall assume that the most desirable arrangement of poles is to have but one multiple pole, and then show that this assumption leads to useful results. Indeed it can be proved at once for the network under consideration that a single multiple pole must lead to the shortest rise time. Thus, the rise time

$$T_R=[2\pi(2-2d_1-d_1^2)]^{\frac{1}{2}},$$

computed from Eq. (26) using Eq. (9), is a minimum when the real, positive coefficient d_1 is maximum. The negative real roots of the denominator of $Z(s)$, $-s_1$, $-s_2$, $\cdots$, $-s_n$, must satisfy the relations

$$\sum(1/s_i)=1$$

and

$$d_1=\sum_{i<j}(1/s_is_j)-\sum_{i<j<k}(1/s_is_js_k)$$

found by multiplying together the n factors

[15] Strictly speaking, two-terminal and three-terminal networks.

[16] See reference 10, pp. 177–181 for a discussion of Foster's reactance theorem and of the various networks which can be used to realize an arbitrary reactance.

$(1+s/s_i)$ of the denominator and comparing the resulting expression with the denominator in Eq. (26). Using Lagrange's method of undetermined multipliers it is found that d_1 is a maximum, i.e., T_R is a minimum when $s_1=s_2=s_3=\cdots$.

The denominator of Eq. (26) can be an even or an odd degree polynomial of degree n, or $n-1$, respectively, depending on whether or not l_0 occurs in the network of Fig. 4. The treatment for both cases follows similar lines, and will be illustrated for the case where $l_0 \neq 0$. In this case we require that $Z(s)$ have one multiple pole of order n (n is always even), and the denominator of Eq. (26), accordingly, must be the binomial expansion of $[1+(s/s_0)]^n$, giving a set of n equations from which the n quantities $s_0, d_1, c_2, d_3, c_4, \cdots, d_{n-1}$ can be determined. The values of the components in the network of Fig. 4 can then be computed, as well as a value for the rise-time figure-of-merit S, and an expression for the transient response to a unit impulse applied at $t=0$.

The computation suggested has been carried out for cases where $Z(s)$ has poles of order $p=1, 2, 3, 4$, and 5, as well as for the limiting case where $p\to\infty$. The following general expressions are found from Eqs. (8) and (9) for the delay time and for the rise-time figure-of-merit

$$\left.\begin{aligned} T_D &= \frac{2}{3}+\frac{1}{3p^2} \\ \text{and} \\ S &= \frac{3}{(2+8/p^2-1/p^4)^{\frac{1}{2}}}, \end{aligned}\right\} \qquad (27)$$

where p is the order of the multiple pole of $Z(s)$. A summary of the results derived from the computations is presented in Table I. The analysis employed for the limiting case is given in Appendix I.

The RC network ($p=1$) has been included in Table I to serve as a basis for comparing the other critically compensated networks. The network for $p=2$ is the well-known, shunt-compensated network, used as an illustration in Section 2. By increasing the value of the inductance from 0.25 to 0.296, and shunting it with a capacitance of 0.125 ($p=3$), a decrease in rise time of about 17 percent is obtained. Adding a second inductance ($p=4$) results in a further decrease in rise time of about 7 percent. By adding more and more components, the limiting value for the rise-time figure-of-merit, $S_{\max}=3/\sqrt{2}=2.12$ is approached. The remaining improvement possible in the transient behavior after a few inductances and capacitances are incorporated in the network is not very marked. These cases, therefore, are not of great practical importance. The limiting case ($p\to\infty$) is of interest primarily because it possesses the greatest figure-of-merit possible with a network of the type under consideration. It is conjectured that this network has the greatest figure-of-merit possible for a low pass two-terminal interstage network. No completely adequate proof, however, has been found for this theorem.

The transient response to a unit step function for all the cases listed in Table I has a monotonic form, which, of course, is necessary in order that the method used for computing delay time and rise time be applicable. The general proof that the transient response is monotonic for arbitrary values of p appears to present considerable algebraic difficulties.

APPENDIX I

Case Where $p\to\infty$

The analysis for the case where the reactance $X(s)$ in Fig. 3b has an infinite number of poles can be made by setting the denominator in Eq. (26) equal to $[1+(s/p)]^p$ and then writing the resulting expression in the algebraically equivalent form

$$Z(s)=1/s-1/2s^2\left[1-\left(1-\frac{s}{p}\Big/1+\frac{s}{p}\right)^p\right].$$

In the limit where $p\to\infty$, this expression becomes

$$Z(s)=1/s-1/2s^2(1-\epsilon^{-2s}) \qquad (28)$$

which has the inverse Laplace transform given in Table I.[17]

From Eq. (28) and the network of Fig. 3b, the reactance $X(s)$ is found to have the form

$$X(s)=\coth s-1/s. \qquad (29)$$

[17] It is of interest to note that if a switch is inserted in series with the capacitor C in Fig. 3b, and the capacitor is initially given a unit charge, then Eq. (28) is the Laplace transform of the voltage developed across the network when the switch is closed at $t=0$. Since the voltage across the capacitor decreases linearly while it is being discharged into the remaining branch of the network, the current flowing through the resistor must have the form of a rectangular pulse (of amplitude $\frac{1}{2}$). The network can evidently be used (ideally, at any rate) to convert either a current impulse, or the sudden discharge of a capacitor, into a rectangular voltage pulse across a resistive load.

The zeros and poles of $X(s)$ are located, of course, on the real frequency axis, and are given by the roots of $\tan\omega=\omega$, and $\sin\omega/\omega=0$, respectively.

To determine values of l_k and c_k, the expression for the reactance, Eq. (29), can be expanded in the infinite series[18]

$$X(s)=\coth s-1/s=\sum_1 \frac{2s}{s^2+k^2\pi^2}. \tag{30}$$

Each term in the infinite series can be interpreted as the reactance of a parallel combination of inductance and capacitance, where

$$\left.\begin{aligned} l_k&=2/\pi^2k^2,\\ c_k&=1/2. \end{aligned}\right\}(k=1, 2, 3, \cdots). \tag{31}$$

It is evident from the nature of the terms occurring in the infinite series that the inductance l_0 must vanish.

The formulas for delay time and rise-time figure-of-merit, Eqs. (27), hold in the limit when $p\to\infty$, so no separate computation need be made for these quantities.

This paper is based on work performed under Contract No. W-7405-Eng-36 with the Manhattan Project at the Los Alamos Scientific Laboratory of the University of California.

[18] See, for instance, E. T. Whittaker and G. N. Watson, *Modern Analysis*, (Cambridge University Press, Tedddington, England, 1927), fourth edition, p. 136, example 7.

A General Divergence Formula*

H. J. Riblet** and C. B. Barker***
Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received April 28, 1947)

A divergence expression for the ratio of energy per steradian reflected from a smooth curved surface to that incident on the surface is derived. It generalizes previous results in that the source and point of observation may both be at finite distances from the reflecting surface. No restrictions are placed on the angles of incidence and reflection except that they be equal. The only limitation placed on the analytical accuracy of the geometrical result is that the surface be sufficiently smooth so that the principal radii of curvature are defined at the point of reflection. It is required, of course, that the wave-length of the energy shall be small compared to the principal radii of curvature of the surface, in order that the geometrical result may be interpreted as a divergence formula. All of the previous results on this problem known to the authors are derived as special cases. Application of the result in connection with the spreading of radio rays by the curvature of the earth leads to somewhat simpler formulas than now available.

INTRODUCTION

A PROBLEM of basic importance in the radio and radar art is that of determining the amount of energy reflected in any direction from a given object placed in a plane or spherical electromagnetic wave. The exact solution of this type of reflection problem is known for only a few special cases. However, if the wave-length is small compared to the radii of curvature of the surface of the reflecting object, it is possible to handle this question by means of geometrical optics. The cases of normal incidence on a general curved surface for a plane and spherical wave have been discussed, respectively, by Goudsmit and Carlson[1] and Silver.[2] Spencer[3] has extended the analysis to arbitrary angles of incidence with the restriction that the wave be plane, while Barker and Riblet[4] have discussed the general case of a spherical wave under the assumption that the point of observation is infinitely far from the reflecting object. This

* This paper is based in part on work done for the Office of Scientific Research and Development under Contract OEMsr-262. It was originally submitted under the title "Reflection from curved surfaces" having Eq. 18 as its principal result. A question of accuracy raised by the reviewer led us to the general result of Eq. 13.

** Now at the Submarine Signal Company, Boston, Massachusetts.

*** Now at Naval Research Laboratory, Washington, District of Columbia.

[1] S. A. Goudsmit and J. F. Carlson, "Microwave radar reflections," Radiation Laboratory Report, 43-13, February 20, 1943.

[2] S. Silver, "Contribution of the dish to the impedance of an antenna," Radiation Laboratory Report, 442, September 17, 1943.

[3] R. C. Spencer, "Reflections from smooth curved surfaces," Radiation Laboratory Report, 661, January 26, 1945.

[4] C. B. Barker and H. J. Riblet, "Reflections from curved surfaces," Radiation Laboratory Report, 976, February 1, 1946.

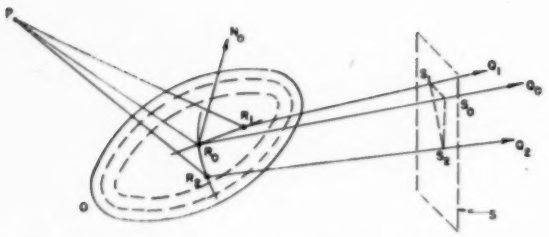


FIG. 1. Geometry of divergence factor calculation.

paper considers the general problem of arbitrary angles of incidence with the source and point of observation both at finite distances. The results mentioned above are all obtained as special cases of the general formula. In the event that the reflecting object is a sphere, our formula is equivalent to an expression which Balth van der Pol and H. Bremmer[5] obtained by studying the exact solution of the diffraction problem in the limit of small wave-lengths.

The need for this type of result was encountered when the writers were confronted with the problem of predicting the radiation patterns that might be expected when microwave antennas of known characteristics were mounted on aircraft. Although pattern measurements on scale models have been very successful at longer wave-lengths, they require scaling the wave-length as well as the plane. Three- and ten-centimeter wave-lengths lie so close to the end of the available frequency spectrum that, at best, only small scale factors are possible. This and the fact that an airplane with a microwave antenna mounted on it is an enormous secondary antenna, which requires large distances for reliable pattern measurements, has made scale model measurements on microwave antennas mounted on aircraft impractical. Fortunately, however, the large ratio of plane size to wave-length leads one to believe that a theoretical approach may be profitable. For an idea of the excellent agreement between experiment and theory and a discussion of the behavior to be expected of certain 3- and 10-centimeter antennas when mounted on aircraft, the reader is referred to the report of reference 4.

Since the derivation of this divergence expression is based on purely geometrical considerations and makes no use of the vector character of the electromagnetic field, it is believed that this result may be of use in other branches of physics where reflections of an "optical" nature are encountered.

In Fig. 1, we consider a source of energy at P and a reflecting object O. If the energy density per steradian incident on O at R_0 is denoted by E_i^2 and the energy density per unit area, at a distance d_2, reflected along the line R_0Q_0 is E_r^2, then we wish to determine E_r^2/E_i^2. It is assumed

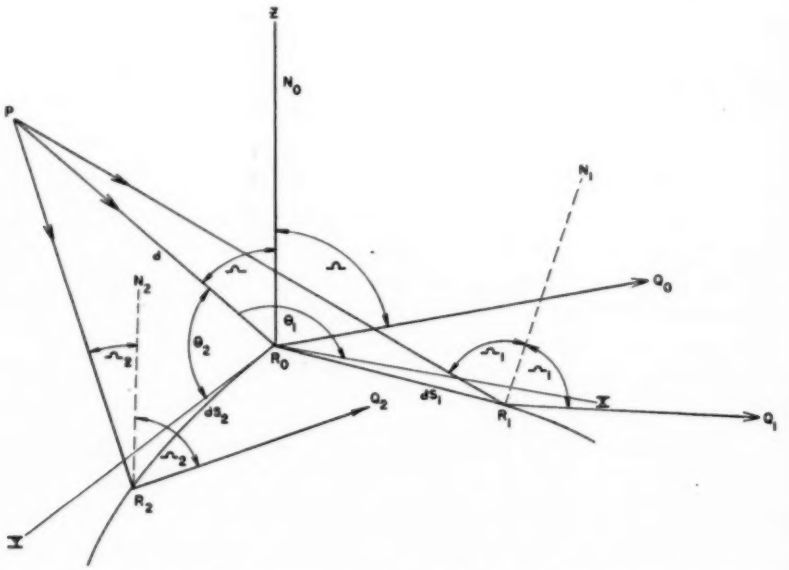


FIG. 2. Geometry at point of reflection.

[5] Balth van der Pol and H. Bremmer, "The propagation of radio waves over a finitely conducting spherical earth with applications to radio telegraphy and the theory of the rainbow," Phil. Mag. **24**, 825; Phil. Mag. **24**, 141. (1937).

that the angle between the line PR_0 and the normal to O at R_0, R_0N_0, is equal to angle $N_0R_0Q_0$. We will also require that the surface be sufficiently smooth so that principal radii of curvature are defined at R_0. The surface may be convex or concave at R_0 as seen from P, however. For a complete discussion of the notion of principal radii of curvature, the reader is referred to Eisenhart.[6] Briefly, they are the reciprocals of the two extreme values of the reciprocal radii of the osculating circles determined by the curves along which planes containing the surface normal R_0N_0 intersect the surface. The tangents to the two selected curves will be referred to as the principal tangents at R_0. It is a very beautiful theorem that these lines are always perpendicular to each other.

The method of solving this problem will be made clear by reference to Fig. 1. Energy from P strikes the object O at two points, R_1 and R_2, near R_0 on the curves through R_0 which determine the principal radii of curvature of the surface at R_0. It is reflected along lines R_1Q_1 and R_2Q_2 in accordance with the requirement that the angles of incidence and reflection be equal. If S is a surface perpendicular to R_0Q_0 at a distance d_2 from R_0, then the reflected rays determine, by intersection, the triangle (S_0, S_1, S_2) whose area will be denoted by ΔS. If E_r^2 is the energy per unit area at a distance d_2 from R_0 measured along R_0Q_0, then the total energy passing through (S_0, S_1, S_2) is $E_r^2\Delta S$. This, however, must equal the energy contained in the spherical angle defined by R_0, R_1, R_2 about point P which we call Δs. Thus

$$E_i^2\Delta s = E_r^2\Delta S, \tag{1}$$

where E_i^2 is the energy per steradian leaving P in the direction R_0. The apparently inconsistent units are chosen because they appear to give the most useful form to our results. We define the divergence factor, D, to be limit of E_r^2/E_i^2 as points R_1 and R_2 approach R_0. In this form it is a function of d_2 as well as the reflecting geometry. From (1)

$$D = \lim_{\Delta s \to 0} \frac{\Delta s}{\Delta S}. \tag{2}$$

[6] L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton University Press, Princeton, New Jersey, 1940), pp. 222–226.

Our principal problem is thus a straightforward exercise in classical differential geometry. Reasonably diligent search and inquiry have failed to disclose this calculation in the mathematical literature so that it is presented here in detail.

EVALUATION OF D

In Fig. 2, we let P be a point source, from which three rays strike a curved surface at R_0, R_1, and R_2. Suppose that R_1 is a point on one of the lines of curvature through R_0, and R_2 lies on the other line of curvature through R_0. Suppose that N_0, N_1, and N_2 are on the normals at R_0, R_1, and R_2, respectively, and let Q_0, Q_1, and Q_2 represent points on the reflected rays from R_0, R_1, and R_2, respectively.

In order to simplify the calculations, let R_0 be the origin of the coordinate system, let the principal tangent in the direction of R_1 be the positive x axis, let the principal tangent in the direction of R_2 be the positive y axis and let the normal at R_0 be the positive z axis. This, as we have already stated, is always possible. Suppose, further, that the direction cosines of the line joining P and R_0 are $\cos\theta_1$, $\cos\theta_2$, and $\cos\Omega$, and that the distance from P to R_0 is d_1. Thus the coordinates of P are ($d_1\cos\theta_1$, $d_1\cos\theta_2$, $d_1\cos\Omega$). The coordinates of R_1, except for differentials of order higher than first, are (ds_1, 0, 0), ds_1 being the arc length measured from R_0 to R_1. Similarly, the coordinates of R_2 are (0, ds_2, 0), ds_2 being the arc length from R_0 to R_2.

Now the direction cosines of PR_1 and PR_2 are, respectively,

$$\frac{d_1\cos\theta_1 - ds_1}{A}, \quad \frac{d_1\cos\theta_2}{A}, \quad \frac{d_1\cos\Omega}{A},$$

and

$$\frac{d_1\cos\theta_1}{B}, \quad \frac{d_1\cos\theta_2 - ds_2}{B}, \quad \frac{d_1\cos\Omega}{B},$$

where

$$A = ((d_1\cos\theta_1 - ds_1)^2 + d_1^2\cos^2\theta_2 + d_1^2\cos^2\Omega)^{\frac{1}{2}},$$

and

$$B = (d_1^2\cos^2\theta_1 + (d_1\cos\theta_2 - ds_2)^2 + d_1^2\cos^2\Omega)^{\frac{1}{2}}.$$

Furthermore, the direction cosines of R_0N_0 are 0, 0, 1 while those of R_1N_1 and R_2N_2 are, respectively, ds_1/ρ_1, 0, $(\rho_1^2 - ds_1^2)^{\frac{1}{2}}/\rho_1$ and 0, ds_2/ρ_2,

$(\rho_2^2-ds_2^2)^{\frac{1}{2}}/\rho_2$, where ρ_1 and ρ_2 are the principal radii of curvature in the direction of R_1 and R_2, respectively.[7]

Suppose l_1, l_2, and l_3 are concurrent, coplanar lines, and that their direction cosines are, respectively, a_i, b_i, and c_i, $(i=1, 2, 3)$. In addition, suppose l_2 bisects the angle between l_1 and l_3. Then, if w is the angle between l_1 and l_2,

$$c_i=2b_i\cos w-a_i \quad (i=1, 2, 3). \tag{3}$$

Now let Ω_1 designate the angle PR_1N_1. Then the angle $N_1R_1Q_1$ is also Ω_1, since the angle of incidence must equal the angle of reflection. By a well-known formula of analytic geometry,

$$\cos\Omega_1=\frac{1}{A\rho_1}(d_1\cos\theta_1 ds_1-ds_1^2 +d_1\cos\Omega(\rho_1^2-ds_1^2)^{\frac{1}{2}}).$$

In the same manner,

$$\cos\Omega_2=\frac{1}{B\rho_2}(d_1\cos\theta_2 ds_2-ds_2^2 +d_1\cos\Omega(\rho_2^2-ds_2^2)^{\frac{1}{2}}),$$

Ω_2 being the angle PR_2N_2.

Now making use of formula (3), the direction cosines of R_1Q_1, say α_1, β_1, γ_1, may be computed. Thus

$$\alpha_1=\frac{2d_1\cos\theta_1 ds_1^2-2ds_1^3+2d_1\cos\Omega(\rho_1^2-ds_1^2)^{\frac{1}{2}}ds_1 -d_1\cos\theta_1\rho_1^2+\rho_1^2 ds_1}{A\rho_1^2},$$

$$\beta_1=\frac{-d_1\cos\theta_2}{A},$$

$$\gamma_1=\frac{2d_1\cos\theta_1(\rho_1^2-ds_1^2)^{\frac{1}{2}}ds_1-2(\rho_1^2-ds_1^2)^{\frac{1}{2}}ds_1^2 -2d_1\cos\Omega ds_1^2+d_1\cos\Omega\rho_1^2}{A\rho_1^2}.$$

Disregarding differentials of order higher than the first, these become:

$$\alpha_1=\frac{2d_1\cos\Omega ds_1-d_1\cos\theta_1\rho_1+\rho_1 ds_1}{A\rho_1},$$

$$\beta_1=\frac{-d_1\cos\theta_2}{A}, \tag{4}$$

$$\gamma_1=\frac{2d_1\cos\theta_1 ds_1+d_1\cos\Omega\rho_1}{A\rho_1}.$$

[7] The simple form which these direction cosines take is the justification for selecting R_1 and R_2 on the lines of curvature.

Similarly, the direction cosines of R_2Q_2, say $(\alpha_2, \beta_2, \gamma_2)$, are:

$$\alpha_2=\frac{-d_1\cos\theta_1}{B},$$

$$\beta_2=\frac{2d_1\cos\Omega ds_2-d_1\cos\theta_2\rho_2+\rho_2 ds_2}{B\rho_2}, \tag{5}$$

$$\gamma_2=\frac{2d_1\cos\theta_2 ds_2+d_1\cos\Omega\rho_2}{B\rho_2}.$$

In addition, the direction cosines of R_0Q_0 are easily found to be

$$\alpha_0=-\cos\theta_1, \quad \beta_0=-\cos\theta_2, \quad \gamma_0=\cos\Omega. \tag{6}$$

In order to determine the area of the triangle (S_0, S_1, S_2), we first determine the coordinates of these three points. The equation of the plane S perpendicular to line R_0Q_0 is

$$\alpha_0x+\beta_0y+\gamma_0z=d_2, \tag{7}$$

where d_2 is the distance from R_0 to S_2. The equations of lines R_0Q_0, R_1Q_1, and R_2Q_2 are, respectively,

$$\frac{x}{\alpha_0}=\frac{y}{\beta_0}=\frac{z}{\gamma_0}, \quad \frac{x-ds_1}{\alpha_1}=\frac{y}{\beta_1}=\frac{z}{\gamma_1}, \tag{8}$$

and

$$\frac{x}{\alpha_2}=\frac{y-ds_2}{\beta_2}=\frac{z}{\gamma_2}.$$

The coordinates of the points S_0, S_1, and S_2 are found, after solving Eqs. (8) successively with (7), to be, respectively,

$$(\alpha_0 d_2, \beta_0 d_2, \gamma_0 d_2),$$

$$\left(\frac{\alpha_1(d_2-\alpha_0 ds_1)}{(\alpha_0, \alpha_1)}+ds_1, \frac{\beta_1(d_2-\alpha_0 ds_1)}{(\alpha_0, \alpha_1)}, \frac{\gamma_1(d_2-\alpha_0 ds_1)}{(\alpha_0, \alpha_1)}\right) \tag{9}$$

and

$$\left(\frac{\alpha_2(d_2-\beta_0 ds_2)}{(\alpha_0, \alpha_2)}, \frac{\beta_2(d_2-\beta_0 ds_2)}{(\alpha_0, \alpha_2)}+ds_2, \frac{\gamma_2(d_2-\beta_0 ds_2)}{(\alpha_0, \alpha_2)}\right),$$

where

$$(\alpha_0, \alpha_1)=\alpha_0\alpha_1+\beta_0\beta_1+\gamma_0\gamma_1,$$

and

$$(\alpha_0, \alpha_2) = \alpha_0\alpha_2 + \beta_0\beta_2 + \gamma_0\gamma_2. \tag{10}$$

If we now make a change of variables which places the origin of the coordinate system at S_0, the coordinates of points S_1 and S_2 are found by suitably subtracting the components of expressions (9). For example, the relative x coordinate, S_{1x}, of S_1 is given by

$$\frac{\alpha_1(d_2-\alpha_0 ds_1)}{(\alpha_0, \alpha_1)} + ds_1 - \alpha_0 d_2$$
$$= \frac{1}{(\alpha_0, \alpha_1)}\{d_2[\alpha_1 - \alpha_0(\alpha_0, \alpha_1)] - \alpha_1\alpha_0 ds_1 + (\alpha_0, \alpha_1)ds_1\},$$

which becomes upon substitution

$$S_{1x} = \frac{ds_1}{(\alpha_0, \alpha_1)A\rho_1}\{2d_1d_2\cos\Omega + \rho_1\sin^2\theta_1(d_1+d_2)\},$$

where we have used

$$(\alpha_0, \alpha_1) = \frac{d_1\rho_1 - \rho_1\cos\theta_1 ds_1}{A\rho_1},$$

and otherwise replaced the direction cosines α_0, $\beta_0\cdots$, by the values given in Eqs. (4)–(6). In a similar manner the relative y and z coordinates of S_1 are found to be,

$$S_{1y} = \frac{ds_1}{(\alpha_0, \alpha_1)A\rho_1}\{-\rho_1\cos\theta_1\cos\theta_2(d_2+d_1)\},$$

and

$$S_{1z} = \frac{ds_1}{(\alpha_0, \alpha_1)A\rho_1}\{2d_1d_2\cos\theta_1 + \rho_1\cos\theta_1 \times\cos\Omega(d_1+d_2)\}.$$

Similarly, the x, y, and z coordinates of S_2 relative to S_0 are, respectively,

$$S_{2x} = \frac{ds_2}{(\alpha_0, \alpha_2)B\rho_2}\{-\rho_2\cos\theta_1\cos\theta_2(d_2+d_1)\},$$

$$S_{2y} = \frac{ds_2}{(\alpha_0, \alpha_2)B\rho_2}\{2d_1d_2\cos\Omega + \rho_2\sin^2\theta_2(d_1+d_2)\},$$

$$S_{2z} = \frac{ds_2}{(\alpha_0, \alpha_2)B\rho_2}\{2d_1d_2\cos\theta_2 + \rho_2\cos\Omega\cos\theta_2(d_1+d_2)\}.$$

If we consider (S_{1x}, S_{1y}, S_{1z}) and (S_{2x}, S_{2y}, S_{2z}) to be the components of two vectors, we can readily see, by reference to the vector product, that

$$4(\Delta S)^2 = (S_{1y}S_{2z} - S_{2y}S_{1z})^2 + (S_{1x}S_{2z} - S_{2x}S_{1z})^2 + (S_{1x}S_{2y} - S_{2x}S_{1y})^2.$$

If we write

$$F = \cos\Omega[4d_1^2d_2^2 + \rho_1\rho_2(d_2+d_1)^2] + (\rho_1\sin^2\theta_1 + \rho_2\sin^2\theta_2)[2d_1d_2^2 + 2d_1^2d_2],$$

we find after elementary calculation that, except for sign possibly,

$$(S_{1y}S_{2z} - S_{2y}S_{1z}) = \frac{\cos\theta_1 F ds_2 ds_1}{(\alpha_0, \alpha_1)(\alpha_0, \alpha_2)AB\rho_1\rho_2},$$

$$(S_{1x}S_{2z} - S_{2x}S_{1z}) = \frac{\cos\theta_2 F ds_2 ds_1}{(\alpha_0, \alpha_1)(\alpha_0, \alpha_2)AB\rho_1\rho_2},$$

and

$$(S_{1x}S_{2y} - S_{2x}S_{1y}) = \frac{\cos\Omega F ds_2 ds_1}{(\alpha_0, \alpha_1)(\alpha_0, \alpha_2)AB\rho_1\rho_2}.$$

Therefore,

$$\Delta S = \frac{F ds_1 ds_2}{2(\alpha_0, \alpha_1)(\alpha_0, \alpha_2)AB\rho_1\rho_2}. \tag{11}$$

Now we have, clearly, that Δs is given by

$$\Delta s = \frac{ds_1 ds_2\cos\Omega}{2d_1^2} \tag{12}$$

so that

$$\frac{\Delta s}{\Delta S} = \frac{\cos\Omega\rho_1\rho_2}{F}\frac{AB(\alpha_0, \alpha_1)(\alpha_0, \alpha_2)}{d_1^2}.$$

In the limit as $\Delta s \to 0$, $A \to B \to d_1$ and $(\alpha_0, \alpha_1) \to (\alpha_0, \alpha_2) \to 1$ so that

$$D = \frac{\cos\Omega\rho_1\rho_2}{F}. \tag{13}$$

It is interesting to observe the symmetry in d_1 and d_2 of expression (13) for the divergence factor. This is, of course, a geometrical analogue of the reciprocity theorem for electromagnetic radiation.

In the event that $d_2 = 0$, we get from (13) that $D = 1/d_1^2$. This is, of course, what we would expect. If R_1 and R_2 are infinite, then

$$D = 1/(d_1+d_2)^2,$$

as one could easily calculate for a plane surface.

SPECIAL CASES

There are a considerable number of special cases in the literature, all of which may be obtained from expression (13). The ratio of the energy density per unit area E_r^2 to the incident energy density per steradian is given, as we have seen, by

$$\frac{E_r^2}{E_i^2}=D=\frac{\cos\Omega\rho_1\rho_2}{F}.$$

In the event that the incident energy is a plane wave we must allow $d_1\to\infty$ so that E_i^2/d_1^2 is a constant equal to the incident energy per unit area. Then

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/unit area}}=\frac{\rho_1\rho_2\cos\Omega}{\cos\Omega[4d_2^2+\rho_1\rho_2]+(\rho_1\sin^2\theta_1+\rho_2\sin^2\theta_2)2d_2}. \quad (14)$$

When the observations are to be made at infinity, we must determine $E_r^2d_2^2$ which is the energy per steradian reflected from the surface. Then letting $d_2\to\infty$ we have

$$\frac{\text{Reflected energy/steradian}}{\text{Incident energy/unit area}}=\frac{\rho_1\rho_2}{4}. \quad (15)$$

This is Spencer's[3] result except for a slight difference in definition. It is independent of the angle of incidence, as Spencer pointed out, and accordingly contains the result used by Goudsmit and Carlson[1] as a special case.

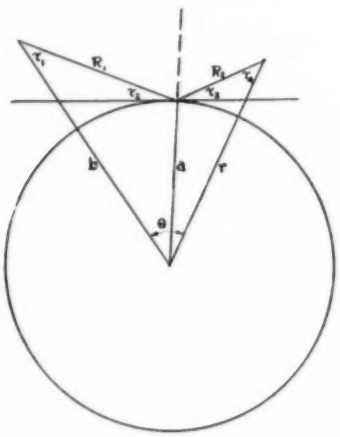


FIG. 3. Geometry for spherical earth calculations.

If the reflecting surface is a cylinder, we obtain our result from (14) by allowing $\rho_1\to\infty$. We then have

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/unit area}}=\frac{\rho_2\cos\Omega}{\cos\Omega\rho_2+2d_2\sin^2\theta_1}. \quad (16)$$

If observations are made at infinity and if $\theta_2=90°$, we are interested in the reflected energy per radian which is $E_r^2d_2$. Thus

$$\frac{\text{Reflected energy/radian}}{\text{Incident energy/unit area}}=\frac{\rho_2\cos\Omega}{2\sin^2\theta_1}. \quad (17)$$

If the incident energy is a spherical wave while the observations are made at infinity, we are interested in the reflected energy per steradian which is $E_r^2d_2^2$. Allowing $d_2\to\infty$ in (13), we have

$$\frac{\text{Reflected energy/steradian}}{\text{Incident energy/steradian}}=\frac{\rho_1\rho_2\cos\Omega}{\cos\Omega(4d_1^2+\rho_1\rho_2)+2d_1(\rho_1\sin^2\theta_1+\rho_2\sin^2\theta_2}. \quad (18)$$

This is the result obtained by Barker and Riblet.[4] If the reflecting object is a cylinder, we let $\rho_2\to\infty$ and obtain

$$\frac{\text{Reflected energy/steradian}}{\text{Incident energy/steradian}}=\frac{\rho_1\cos\Omega}{\cos\Omega\rho_1+2d_1\sin^2\theta_2}. \quad (19)$$

When the source and point of observation are both at a finite distance from the reflecting object, there are a number of special cases of interest. For normal incidence $\Omega=0$ and θ_1 and θ_2 equal 90°. Then

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/steradian}}=\frac{\rho_1\rho_2}{\rho_1\rho_2(d_2+d_1)^2+(\rho_1+\rho_2)2d_1d_2(d_1+d_2)+4d_1^2d_2^2}=\frac{\rho_1}{\rho_1(d_2+d_1)+2d_1d_2}\cdot\frac{\rho_2}{\rho_2(d_2+d_1)+2d_1d_2}. \quad (20)$$

If $d_1=d_2=d$ we have

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/steradian}}=\frac{\rho_1\rho_2}{4d^2(\rho_1+d)(\rho_2+d)}. \quad (21)$$

This is the formula obtained by Silver.[2] The apparent difference in sign is a consequence of the fact that we have taken the source to be on the convex side of the surface. This point serves to emphasize the fact that our results are not limited to convex surfaces and tells us that for concave surfaces we have only to make suitable changes of sign in the radii of curvature.

If the reflecting surface is a cylinder, we let $\rho_1\to\infty$ and have

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/steradian}}=\frac{\rho_2\cos\Omega}{\cos\Omega\rho_2(d_1+d_2)^2+\sin^2\theta_1 2d_1d_2(d_1+d_2)}. \quad (22)$$

For a number of reasons, the case when the reflecting object is a sphere is of particular interest. For this purpose we put $\rho_1=\rho_2=R$ in (13) and we have

$$\frac{\text{Reflected energy/unit area}}{\text{Incident energy/steradian}}=\frac{R_2\cos\Omega}{\cos\Omega\{4d_1^2d_2^2+R^2(d_1+d_2)^2\}+R(1+\cos^2\Omega)\cdot 2d_1d_2(d_1+d_2)}$$

$$=\frac{R^2\cos\Omega}{[R(d_1+d_2)+d_1d_2\cos\Omega][2d_1d_2+R\cos\Omega(d_1+d_2)]}. \quad (23)$$

The solution of the same problem is arrived at in two ways by Balth van der Pol and H. Bremmer in one of their articles on the diffraction of electromagnetic waves by a finitely conducting earth.[8] In their notation

$$\alpha_{11}=\frac{a(R_1+R_2)(\sin\tau_2\cos\tau_2)^{\frac{1}{2}}}{(br\sin\theta(R_1r\cos\tau_4+R_2b\cos\tau_1))^{\frac{1}{2}}}. \quad (24)$$

The meaning of the symbols on the right-hand side of this expression is made clear in Fig. 3. By definition

$$(d_1+d_2)^2D=(\alpha_{11})^2,$$

where $d_1=R_1$, $d_2=R_2$, and $\tau_2=\Omega$. It is not difficult to show that these expressions are indeed equivalent. Equation (23) may be somewhat simpler since it is expressed in terms of the minimum number of independent parameters required to characterize the geometry of the problem.

If we consider $D^{-1}(d_1+d_2)^{-2}$ we have

$$1+\frac{2d_1d_2(1+\cos^2\Omega)}{(d_1+d_2)R\cos\Omega}+\frac{4d_1^2d_2^2}{R^2(d_1+d_2)^2}.$$

The error in neglecting the last term is certainly less than 1 percent if d_1 and d_2 are one-tenth of R or less. If the reflecting sphere is the earth and d_1 and d_2 are less than 400 miles, we have, with an accuracy better than 1 percent,

$$D^{-1}(d_1+d_2)^{-2}=1+\frac{2d_1d_2}{R(d_1+d_2)\dfrac{\cos\Omega}{1+\cos^2\Omega}}. \quad (25)$$

To include the effect of the refraction of the rays by the atmosphere, we have only to recalculate d_1, d_2, and Ω, fixing the heights of the source and point of observation and their angular separation but increasing R to KR, where K is 4/3. Expression (25) is what Norton[9] has called D^2. If we replace the d's used by Norton by δ's, we have, on the assumption that $\sin\delta/R=\delta/R$, that

$$d_1=\frac{\delta_1}{\cos\psi_2'-\sin\psi_2'(\delta_1/R)},$$

and

$$d_2=\frac{\delta_2}{\cos\psi_2'-\sin\psi_2'(\delta_2/R)}.$$

Then (25) reduces to Norton's value of D^2 when ψ_2' is small. Expression (25) is fully as simple as Norton's and is not subject to this restriction.

[8] Balth van der Pol and H. Bremmer, see reference 5.

[9] K. A. Norton, "The calculation of ground-wave field intensity over a finitely conducting spherical earth," Proc. I.R.E. 29, 623 (1941), Eq. (24).

CONCLUSION

A general divergence expression is derived relating the reflected energy density to the energy incident on a smooth reflecting surface. This formula exactly solves the reflection problem in the limit as the wave-length approaches zero. It will be an excellent approximation as long as the wave-length is sufficiently small compared to the radii of curvature of the surface. The fundamental question, as to just how small this ratio has to be, is left unanswered. However, the machinery required for an answer is contained in a series of papers by Balth van der Pol and H. Bremmer.[5] Here formula (24) is obtained from the exact solution of the diffraction problem by going to the limit of small wave-lengths. Comparison of the exact expression obtained by these writers with approximate expression involving (24) will enable one to determine the limits of usefulness of formula (13).

It should also be observed that the usefulness of (13) is not limited to perfectly reflecting surfaces. The introduction of a reflection coefficient characteristic of the reflecting material will allow one to apply our results to finitely conducting surfaces. An explanation of this procedure is contained in the papers of reference 5.

The Theory of the Study of Transfers within a Multi-Compartment System Using Isotopic Tracers*

C. W. SHEPPARD**
The Department of Biochemistry, Vanderbilt University, School of Medicine, Nashville, Tennessee
(Received June 18, 1947)

Expressions are derived for a system of n compartments whereby the amount of a substance S which has moved from any one compartment to any other can be determined in a single experiment from measurements of the amounts of n tagged species in the individual compartments. It is shown that n individual species of S are required which must be uniformly mixed and behave identically but must be distinguishable by the observer. The application of the equations to studies using multiple isotopic tracers is described, and the special case of a two compartment system is analyzed in detail.

THERE are numerous instances in biological and chemical research where multiple compartment systems are encountered. This is undoubtedly true in other fields as well. In such a system real compartments may exist whose contents are homogeneous and which are separated from one another by real boundaries. However, the concept may be generalized so that a substance, such as a chemical element, can be considered to be in a different "compartment" when it is in a different state of chemical combination.

It will be shown that, if several species of the substance are available, the amount of the substance which has moved from any one compartment to any other can be determined in a single experiment from observations of the contents of the compartments. Not only will the required number of these species be determined but also it will be of interest to see mathematically how many observations are required and what must be observed.

NUMBER OF SPECIES REQUIRED

Given a system of n compartments containing varying amounts of a substance, S, let the amount in compartment j be S_j. Its change in an interval of time dt will be dS_j. Let the amount taken from compartment k and delivered to compartment j be dS_{jk}. Since the total amount of S is constant,

$$\sum_{j=1}^{j=n} dS_j = 0. \tag{1}$$

* This research was supported by the Bristol-Myers fund.
** Present address: Biology Division, Clinton Laboratories, Oak Ridge, Tennessee.

By definition

$$dS_j=\sum_{k=1}^{k=n}(dS_{jk}-dS_{kj}) \text{ where } j\neq k. \quad (2)$$

If it is desired to determine the dS_{jk}'s from the dS_j's, $n(n-1)$ independent equations are required, since there are that many unknowns. There are n equations in set 2. Only $n-1$ of these are independent because of the relation 1 between the dS_j's. Thus, since there are too few equations, it is not possible to determine from observations only of the changes in the contents of each of the n compartments the amounts which moved from each of them to each of the others. (If all but $n-1$ of the dS_{jk}'s are known, e.g., $=0$, then the remaining $n-1$ might be found.)

Suppose now that the system contains m different varieties of substance S, namely, ${}^1S$, ${}^2S$, ${}^3S$, $\cdots {}^mS$. These varieties are uniformly mixed in any compartment and *behave identically* in the system but are distinguishable by the observer. Instead of 1 and 2 we now have the relations

$$\sum_{j=1}^{j=n} d^rS_j=0 \text{ (one for each value of } r), \quad (3)$$

$$d^rS_j=\sum_{k=1}^{k=n}(d^rS_{jk}-d^rS_{kj}) \text{ (one for each value of } r \text{ and } j). \quad (4)$$

Because of the identical behavior of the ${}^rS$'s and the uniform mixing the fraction of a given species present in the outflow from a given compartment is the same as the fraction present in the compartment; thus the following relation exists:

$$d^rS_{jk}/dS_{jk}={}^rS_k/S_k={}^rA_k. \quad (5)$$

This set of equations defines ${}^rA_k$ as the fraction of variety r in compartment k. Also

$$dS_j=\sum_{r=1}^{r=m} d^rS_j, \quad (6)$$

and

$$dS_{jk}=\sum_{r=1}^{r=m} d^rS_{jk}. \quad (7)$$

Equations in sets 1 and 2 are derived directly from sets 3, 4, 6, and 7 and thus do not constitute any additional independent sets of relations. By combining equations in sets 3, 4, 5, and 7 we obtain the final relations:

$$\sum_{j=1}^{j=n} d^rS_j=0, \quad (3)$$

$$d^rS_j=\sum_{k=1}^{k=n}({}^rA_k dS_{jk}-{}^rA_j dS_{kj}), \quad (8)$$

$$\sum_{r=1}^{r=m} {}^rA_j=1. \quad (9)$$

For each value of r there are n equations in set 8. Of these it can easily be shown that, subject to relation 3, only $n-1$ are independent. We may therefore conclude that $m(n-1)$ must be equal to the number of unknowns $n(n-1)$ in order that the set be soluble. Thus if $m=n$ the necessary condition is fulfilled that the dS_{jk}'s may be determined in terms of the measurable d^rS_j's and the measurable ${}^rA_j$'s. Although this is a necessary condition for solubility, the fact that it is sufficient must rest upon the fact that solutions can be obtained which satisfy the equations.

SOLUTION OF THE EQUATIONS

The solution of the equations is simplified somewhat if a particular system is used in writing them. The case where $n=3$ will illustrate the procedure in general. Thus

$$\begin{array}{cccccc}
-{}^1A_1dS_{31}+{}^1A_2dS_{12}+{}^1A_3dS_{13}-{}^1A_1dS_{21} & 0 & 0 & =d^1S_1,\\
-{}^2A_1dS_{31}+{}^2A_2dS_{12}+{}^2A_3dS_{13}-{}^2A_1dS_{21} & 0 & 0 & =d^2S_1,\\
-{}^3A_1dS_{31}+{}^3A_2dS_{12}+{}^3A_3dS_{13}-{}^3A_1dS_{21} & 0 & 0 & =d^3S_1,\\
0 & -{}^1A_2dS_{12} & 0 & +{}^1A_1dS_{21}-{}^1A_2dS_{32}+{}^1A_3dS_{23}=d^1S_2,\\
0 & -{}^2A_2dS_{12} & 0 & +{}^2A_1dS_{21}-{}^2A_2dS_{32}+{}^2A_3dS_{23}=d^2S_2,\\
0 & -{}^3A_2dS_{12} & 0 & +{}^3A_1dS_{21}-{}^3A_2dS_{32}+{}^3A_3dS_{23}=d_3S_2.
\end{array}$$

The equations are arranged in groups in each of which all indices are the same except for the superscript r which takes on successive values from 1 to n. Each of these groups is characterized by the value of the subscript j of the term on the right side, which varies from 1 to $n-1$. In the arrangement of the terms within the equations a division into groups is also made. Each group

corresponds to a value of j which is the first subscript in each case. Within the group successive terms are characterized by values of the second subscript k, which varies from 1 to n. In the place which would be occupied by the non-existent term for which $j=k$ the terms dS_{n1}, dS_{n2}, etc., are inserted.

The equations are solved by the usual method of determinants. The determinant of the coefficients may be reduced to a particularly simple form which is illustrated for the case where $n=3$, although the general case presents no additional features. Here we have a compound diagonal determinant containing subdeterminants of the form

$$\Delta = -\begin{vmatrix} {}^1A_1 & {}^1A_2 & {}^1A_3 \\ {}^2A_1 & {}^2A_2 & {}^2A_3 \\ {}^3A_1 & {}^3A_2 & {}^3A_3 \end{vmatrix}.$$

The diagonal expressions are flanked by foreign elements containing columns of A's. However, for each of these there is a corresponding column containing the same terms where the rest of the elements of the column are zero. By subtracting the latter column in each instance, the interfering element may be removed. The compound determinant then becomes Δ^2. In the general case the determinant is Δ^{n-1} where Δ contains n rows and columns, the superscripts of the A's denoting the row and the subscripts the column. The determinant of the numerator of a desired dS_{jk} is obtained from the previous determinant in the usual manner by replacing the entire row of coefficients of dS_{jk} by the entire row of the d^rS_j's. For the case where $n=3$ the result for dS_{23} is

$$\begin{vmatrix} -{}^1A_1 & {}^1A_2 & {}^1A_3 & -{}^1A_1 & 0 & d^1S_1 \\ -{}^2A_1 & {}^2A_2 & {}^2A_3 & -{}^2A_1 & 0 & d^2S_1 \\ -{}^3A_1 & {}^3A_2 & {}^3A_3 & -{}^3A_1 & 0 & d^3S_1 \\ 0 & -{}^1A_2 & 0 & {}^1A_1 & -{}^1A_2 & d^1S_2 \\ 0 & -{}^2A_2 & 0 & {}^2A_1 & -{}^2A_2 & d^2S_2 \\ 0 & -{}^3A_2 & 0 & {}^3A_1 & -{}^3A_2 & d^3S_2 \end{vmatrix}.$$

The interfering A's can be removed as before. By a Laplace expansion it can be shown that all of the d^rS_j's which are outside of the diagonal subdeterminants (in this case where $j=2$) can be ignored. Thus the final result is

$$\begin{vmatrix} {}^1A_1 & {}^1A_2 & d^1S_2 \\ {}^2A_1 & {}^2A_2 & d^2S_2 \\ {}^3A_1 & {}^3A_2 & d^3S_2 \end{vmatrix} \times \Delta.$$

The analysis is not valid for the case where $j=n$ (in this case 3). It is of interest that one column of A's always has a negative sign. This is in all cases the column where $j=n$ and is thus never replaced by the d^rS_j's. As a result, both numerator and denominator determinants have the same sign where n is odd or even. We therefore use positive signs throughout in the final result which is

$$dS_{23} = \frac{\begin{vmatrix} {}^1A_1 & {}^1A_2 & d^1S_2 \\ {}^2A_1 & {}^2A_2 & d^2S_2 \\ {}^3A_1 & {}^3A_2 & d^3S_2 \end{vmatrix}}{\begin{vmatrix} {}^1A_1 & {}^1A_2 & {}^1A_3 \\ {}^2A_1 & {}^2A_2 & {}^2A_3 \\ {}^3A_1 & {}^3A_2 & {}^3A_3 \end{vmatrix}}.$$

In the general case

$$dS_{jk} = \frac{\begin{vmatrix} {}^1A_1 & \cdots & {}^1A_{k-1} & d^1S_j & {}^1A_{k+1} & \cdots & {}^1A_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ {}^nA_1 & \cdots & {}^nA_{k-1} & d^nS_j & {}^nA_{k+1} & \cdots & {}^nA_n \end{vmatrix}}{\begin{vmatrix} {}^1A_1 & \cdot & \cdot & \cdot & \cdot & \cdot & {}^1A_n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ {}^nA_1 & \cdot & \cdot & \cdot & \cdot & \cdot & {}^nA_n \end{vmatrix}}. \quad (10)$$

Although the above considerations fail when $j=n$ the final expression is the same in all cases. This can be verified by interchanging j and k throughout and using the expression for the case where $k=n$.

Three conclusions become evident from this result. Firstly, observational access is required to all compartments but only to determine the A's. The second is that direct determination of the changes d^rS_j are required in only one compartment, namely, that into which dS_{jk} is being delivered. Finally, all n of the tracers must be measured.

The solution of the equations fails increasingly as the determinant of the ${}^rA_k$'s approaches zero. A few instances may be cited. In the first, the concentration of a given species is vanishingly small in all compartments. Here all of the terms approach zero in a given row. This situation arises in the particular case where one of the tagged species is eliminated and thus sufficient tags are not provided. A similar situation occurs if the A's for any two or more different species are the same or bear a constant ratio in each compartment. This condition is approached late in the course of any experiment in which all species

tend to become uniformly mixed. If all A's are zero in a given compartment or equal or proportional in any two or more compartments the determinant is also zero. However, in this case the solution for a system of fewer compartments should be investigated.

An interesting special case arises if each compartment contains a single different species of S. Under this condition a given compartment carries its own label, and the amount of this species appearing in another compartment gives the observer the desired information. The analysis is simplified if the superscript denoting the species is chosen to be the same as the number k of the compartment which the species is labeling. Under these conditions ${}^{r}A_k=1$ if $r=k$ and zero otherwise. The determinant Δ contains only diagonal unit terms and thus $\Delta=1$, while $dS_{jk}=d^{r}S_j=d^{k}S_j$. Although this arrangement of tracers simplifies the analysis the difficulty of preparing such a system is obvious and once the transfer among the compartments has progressed the original special situation is rapidly abolished. In the general case the expression which has been derived for dS_{jk} applies at any stage of a given experiment and imposes no restriction on the percentages of the individual species present in the compartments provided that the relatively simple criterion $\Delta\neq 0$ is satisfied.

THE TWO COMPARTMENT SYSTEM

The general problem is illustrated by the special case of a two compartment system. Here $n=m=2$. The Eqs. (3), (8), and (9) have the form

$$d^{1}S_1+d^{1}S_2=0, \quad d^{2}S_1+d^{2}S_2=0, \tag{3'}$$

$$\left.\begin{aligned} d^{1}S_1&={}^{1}A_2dS_{12}-{}^{1}A_1dS_{21}\\ d^{1}S_2&={}^{1}A_1dS_{21}-{}^{1}A_2dS_{12}\\ d^{2}S_1&={}^{2}A_2dS_{12}-{}^{2}A_1dS_{21}\\ d^{2}S_2&={}^{2}A_1dS_{21}-{}^{2}A_2dS_{12}\end{aligned}\right\}, \tag{8'}$$

$${}^{1}A_1-{}^{2}A_1={}^{1}A_2-{}^{2}A_2=1. \tag{9'}$$

Of these, the first and second pairs of Eqs. (8') are related through Eqs. (3'). Thus, they are not independent relations. However, one from each pair can be solved simultaneously, yielding

$$\begin{aligned} dS_{12}&=\frac{{}^{1}A_1d^{2}S_1-{}^{2}A_1d^{1}S_1}{{}^{1}A_1{}^{2}A_2-{}^{1}A_2{}^{2}A_1},\\ dS_{21}&=\frac{{}^{2}A_2d^{1}S_2-{}^{1}A_2d^{2}S_2}{{}^{1}A_1{}^{2}A_2-{}^{1}A_2{}^{2}A_1}.\end{aligned} \tag{10'}$$

These final relations are also obtained from Eq. (10) for the case $n=2$. As should be expected the value of dS_{21} is obtained from dS_{12} by interchanging the subscripts. From the relations between the dS_k's and the A's alternative expressions are obtained such as

$$dS_{21}=\frac{{}^{1}A_2d^{2}S_1-{}^{2}A_2d^{1}S_1}{{}^{1}A_1{}^{2}A_2-{}^{1}A_2{}^{2}A_1}, \text{ etc.}$$

However, it is to be noted that in no case is it possible to obtain expressions in terms of A's measured in one compartment alone. Also, in no case may an expression be obtained in which a single $d^{1}S$ or $d^{2}S$ appears alone. Thus, we see that two species of S are essential and that we must have observational access to both compartments.

THE SPECIFIC ACTIVITIES

Up to this point only those experiments have been considered in which it is possible directly to determine both the relative amount ${}^{r}A_k$ of a given species and the change in absolute amount $d^{r}S_k$ in a given compartment. This may be possible in experiments with stable isotopes where such values are obtained from mass spectrometric analyses. However, important instances arise where these cannot be obtained directly. Of particular interest are experiments in which radioactive isotopes are employed. Here, the actual percentage of a given radioactive isotope (usually very minute) is of no importance. Instead the concentrations of radioactivity ${}^{r}R_k$ (usually in counts per minute per cc) in the various compartments are determined. These quantities are only proportional to ${}^{r}A_k$ and $d^{r}S_k$. In order to demonstrate why it is still possible to obtain dS_{jk} a property of determinants must be cited. Consider the quotient of two determinants, d and D. If all members of corresponding rows of d and D be multiplied by arbitrary constants, which may be different for each pair or rows, the quotient is unaltered. Thus,

$$d/D=\frac{\begin{vmatrix} {}^1d_1 & \cdots & {}^1d_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ {}^nd_1 & \cdots & {}^nd_n \end{vmatrix} \times \begin{vmatrix} \alpha 0 & \cdots & 0 \\ 0\beta & \cdots & 0 \\ \cdot & \cdots & \cdot \\ 00 & \cdots & \omega \end{vmatrix}}{\begin{vmatrix} {}^1D_1 & \cdots & {}^1D_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ {}^nD_1 & \cdots & {}^nD_n \end{vmatrix} \times \begin{vmatrix} \alpha 0 & \cdots & 0 \\ 0\beta & \cdots & 0 \\ \cdot & \cdots & \cdot \\ 00 & \cdots & \omega \end{vmatrix}} = \frac{\begin{vmatrix} \alpha {}^1d_1 & \cdots & \alpha {}^1d_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \omega {}^nd_1 & \cdots & \omega {}^nd_n \end{vmatrix}}{\begin{vmatrix} \alpha {}^1D_1 & \cdots & \alpha {}^1D_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \omega {}^nD_1 & \cdots & \omega {}^nD_n \end{vmatrix}}. \tag{11}$$

If it be assumed that measurements of individual radioactivities are possible, then in place of ${}^rA_k$ the ratio ${}^rR_k/C_k$ must be obtained, where C_k is the total concentration of S in compartment k (usually in milliequivalents or millimols per cc). This ratio is called the specific activity ${}^ra_k$ of species r in compartment k. In all cases, ${}^ra_k$ is proportional to ${}^rA_k$. In place of d^rS_k or $d({}^rC_kV_k)$ where ${}^rC_k$ is the concentration of species r in compartment k, the quantity $d({}^rR_kV_k)$ is required. These two quantities are in the same ratio as ${}^rA_k$ and ${}^ra_k$. In conclusion, it is seen that we can replace every ${}^rA_k$ in Eqs. (10) by ${}^ra_k$ and every d^rS_k by $d({}^rR_kV_k)$. Thus,

$$dS_{jk}=\frac{\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_{k-1} & d({}^1R_jV_j) & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ {}^na_1 & \cdots & {}^na_{k-1} & d({}^nR_jV_j) & {}^na_{k+1} & \cdots & {}^na_n \end{vmatrix}}{\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ {}^na_1 & \cdots & {}^na_n \end{vmatrix}}. \tag{12}$$

It may be concluded that, as before, access to all compartments is required but only to determine the specific activities. In addition, determinations of the change in the total radioactivity of each species is required in compartment j to which dS_{jk} is being delivered. We note that in general $d({}^rR_jV_j)$ contains the two terms ${}^rR_jdV_j$ and $V_jd^rR_j$. The numerator of expression (12) can be further simplified, if it be rewritten as follows:

$$\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_j & \cdots & {}^1a_{k-1} & {}^1R_jdV_j+V_jd{}^1R_j & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ {}^na_1 & \cdots & {}^na_j & \cdots & {}^na_{k-1} & {}^nR_jdV_j+V_jd{}^nR_j & {}^na_{k+1} & \cdots & {}^na_n \end{vmatrix}.$$

$$=\begin{vmatrix} {}^1a_1 & \cdots & {}^1R_j/C_j & \cdots & {}^1a_{k-1} & {}^1R_jdV_j+V_jd{}^1R_j & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ {}^na_1 & \cdots & {}^nR_j/C_j & \cdots & {}^na_{k-1} & {}^nR_jdV_j+V_jd{}^nR_j & {}^na_{k+1} & \cdots & {}^na_n \end{vmatrix}$$

$$=\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_{k-1} & V_jd{}^1R_j & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ {}^na_1 & \cdots & {}^na_{k-1} & V_jd{}^nR_j & {}^na_{k+1} & \cdots & {}^na_n \end{vmatrix},$$

since by a well-known rule we can subtract from the terms of any row (here the k'th), the products of the corresponding terms of any other row (namely, the j'th) by any constant factor, i.e., $C_j dV_j$. Thus, we obtain the interesting conclusion that observation only of the volume of compartment j is required, and not of its change.

Non-Radioactive Species

One of the species of S (i.e. the nth) may be distinguished by being non-radioactive. In this case we replace the last rows of the determinants in Eqs. (10) by the sums of all rows and invoke Eqs. (6) and (9). We apply the transformation used in obtaining Eqs. (12) to all but the last rows. Again all terms containing dV_j drop out, thus:

$$dS_{jk}=\frac{\begin{vmatrix} {}^1a_1 & \cdots & {}^1a_{k-1} & V_j d{}^1R_j & {}^1a_{k+1} & \cdots & {}^1a_n \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & \cdots & 1 & V_j dC_j & 1 & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} {}^1a_1 & \cdots & \cdots & \cdots & \cdots & \cdots & {}^1a_n \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdots & \cdot \\ 1 & \cdots & \cdots & \cdots & \cdots & \cdots & 1 \end{vmatrix}}. \quad (13)$$

This is the practical relation which will ordinarily be used in experiments with radioactive isotopes, since one of the species may always be distinguished by being non-radioactive. The measurements required may be summarized:

1. Radioactivities (in counts per minute per cc) of $n-1$ different radioactive isotopes in all compartments but the jth,
2. Concentrations of S (milliequivalents or millimoles per cc) in all compartments but the jth,
3. Volume (in cc) of compartment j,
4. Change of the $n-1$ radioactivities in compartment j,
5. Change in the concentration of S in compartment j.

Single Radioactive Isotope

A two compartment system requires a single radioactive isotope only. Here, no superscript is needed. Inserting R_1 and R_2 in place of ${}^1R_1$ and ${}^1R_2$ in Eqs. (13) for the case $n=2$ we obtain:

$$dS_{12}=V_1\frac{(a_1 dC_1-dR_1)}{a_1-a_2},$$

and

$$dS_{21}=V_2\frac{(a_2 dC_2-dR_2)}{a_2-a_1}.$$

As before $a_1=R_1/C_1$, etc.

INTEGRATION OF DIFFERENTIAL EXPRESSIONS

The above expressions must be integrated if they are to be used in determining the finite transfers ΔS_{jk} which occur during the course ot a given experiment. The general procedure is besf indicated by integrating the expressions for the two compartment case, thus:

$$\Delta S_{12}=\int_{C_1}^{C_2}\frac{a_1 V_1 dC_1}{a_1-a_2}-\int_{R_1}^{R_2}\frac{V_1 dR_1}{a_1-a_2}$$

$$=\overline{a_1 V_1/(a_1-a_2)}\Delta C_1-\overline{V_1/(a_1-a_2)}\Delta R_1$$

where the expressions under the first and second bars are averaged over C_1 and R_1, respectively, as independent variables. When V_1 is constant and a_1 and a_2 vary linearly with S_{12}, C_1, and R_1 (in particular when all vary linearly with time) a simple relation is obtained:

$$\overline{a_1-a_2}\Delta S_{12}=V_1(\bar{a}_1\Delta C_1-\Delta R_1),$$

thus

$$\Delta S_{12}=V_1\frac{(a_1'\Delta C_1-\Delta R_1)}{a_1'-a_2'},$$

where the primes represent values taken halfway through the interval. Similarly,

$$\Delta S_{21}=V_2\frac{(a_2'\Delta C_2-\Delta R_2)}{a_2'-a_1'}.$$

MULTIPLE EXPERIMENTS

If the values of dS_{jk} are to be obtained in a single experiment it is necessary to employ n individual species of S which are distinguishable one from the other on the basis of different (generally physical) properties. However, there are sometimes too few species available. Thus,

in the case of the important biological element phosphorus, only two isotopes, P^{31} and P^{32}, can be used in practice. This difficulty can be overcome only if it is possible to perform the same experiment more than once. Under these conditions the same species can be used a second time or repeatedly until sufficient measurements are achieved. Physically identical species now become different because they belong to different experiments. In the simplest multi-compartment system, $n=2$. Thus, at least two varieties of S will always be required. Using only two species in a four compartment system we must perform two experiments in order to obtain the necessary four separate ${}^{r}A_k$ values in each compartment. Where $n=3$, two experiments will also be needed, but only three of the four ${}^{r}A_k$'s will yield independent information. In general, if m species are available and p experiments are performed, mp must be at least equal to n.

Although the experiments must be identical in that the total amount of S in each compartment must be the same at any stage of each experiment, the percentages of the individual species used for tagging must be different at each repetition since otherwise one or more rows of Δ will be identical and this determinant will vanish. Thus, only if the ${}^{r}A_k$'s are different in a second experiment can this experiment be distinguished from the first. The biologist will readily appreciate that at times the repetition of an experiment under identical conditions is a rather difficult undertaking.

DISCUSSION

It is readily apparent that even in a system of only three compartments, a large number of measurements must be made in order to determine the desired quantities. As the number of compartments increases the method becomes increasingly cumbersome. If rapid variations occur in the various observed quantities (such as the a's for example) many points must be taken to properly describe these variations, and in the final averaging graphical integrations may often be required. However, in many instances simplifications will arise. Thus in some biological systems, the total quantity of S in one or all compartments may be held constant, since otherwise the system would be disturbed from its normal state. Often the volumes of the compartments will be constant and in many instances observations may be made which are sufficiently close together that linear variations of all quantities may be correctly assumed.

In the use of radioactive tracers different types of radioactivity must be available which may be separately distinguished. If half-lives differ sufficiently (e.g., Na^{22} of half-life 3.0 years and Na^{24} of half-life 14.5 hours) then the sum of the two activities may be determined by an early measurement and later, the activity of the long-lived member may be measured when the short period activity has declined to a negligible level. Methods are also available in which instruments are used which respond to different types of radiation. Counters have been constructed with which the soft K capture x-rays of Fe^{55} can be determined in the presence of the β rays from Fe^{59} and vice versa.[1]

It must not be assumed that only isotopic experiments are to be analyzed by the equations derived in the present discussion. Of occasional importance are such problems as the motion of fluids in a hydraulic system where dyes are used which remain truly dissolved and thus satisfy the criterion of indistinguishable behavior. However, the above analysis should be of principal importance in the field of radioactive and stable isotopes.

[1] W. C. Peacock and W. M. Good, Rev. Sci. Inst. **17**, 255 (1946).

On the Stresses in a Plate Containing Two Circular Holes

CHIH-BING LING
Guggenheim Laboratory, California Institute of Technology, Pasadena 4, California
(Received May 29, 1947)

This note gives a theoretical solution to a plate containing two circular holes of equal size. The method of solution is to add to the given stress system a suitable biharmonic function which gives no stress at infinity. The parametric coefficients involved in the solution are adjusted so as to satisfy the boundary conditions at the edges of the holes. Bipolar coordinates are used in the solution, by means of which explicit expressions are obtained for the parametric coefficients. Three fundamental stress systems are discussed in some detail, namely, the all-around tension case, the longitudinal tension case, and the transverse tension case. Formulas of the stress along the edges of holes are derived and, in particular, values of maximum stress are calculated. The limiting case in which the holes are tangential is also discussed.

THE general solution of the stresses in a perforated plate, of infinite size, containing a group of circular holes was obtained by Howland and Knight,[1] and Green.[2] Naturally, such general solutions do not give much information as to the numerical aspect of the problem unless particular cases are worked out separately. In both solutions, it may be mentioned that the parametric coefficients involved in the stress function are solved by the method of successive approximations, which is, however, a lengthly process.

The simplest case of a perforated plate is the one containing a single circular hole. This problem has been discussed in great detail by various investigators. The next less simple case is a plate containing a group of two circular holes of equal size. Owing to its importance in engineering applications, it is thought that it is also necessary to discuss this problem in some detail.

A plate problem involving two circular boundaries is best treated by means of bipolar coordinates. A general discussion of stress and strain in bipolar coordinates was given in a paper by Jeffery sometime ago.[3] Nevertheless, the problem of a plate containing two equal holes has not been worked out as fully as it deserves. In this note, a solution of this problem, differed slightly from Jeffery's, will be given, corresponding to several fundamental stress systems acting in the plate. The parametric coefficients involved in the solution are obtained in explicit forms in terms of bipolar coordinates. Finally, the maximum stresses in the plate are calculated and plotted against the distance between the holes.

The bipolar coordinates (ξ, η) will be defined by the equation of transformation

$$x+iy=-a\coth\tfrac{1}{2}i(\xi+i\eta), \tag{1}$$

such that the two poles of the coordinates are located on the x axis at the points $(\pm a, 0)$, and

$$x=J\sinh\eta, \quad y=J\sin\xi, \tag{2}$$

where

$$a/J=\cosh\eta-\cos\xi. \tag{3}$$

The biharmonic equation $\nabla^4\chi=0$, which must be satisfied by the stress function χ, then transforms to

$$\left(\frac{\partial^4}{\partial\xi^4}+2\frac{\partial^4}{\partial\xi^2\partial\eta^2}+\frac{\partial^4}{\partial\eta^4}+2\frac{\partial^2}{\partial\xi^2}-2\frac{\partial^2}{\partial\eta^2}+1\right)\frac{\chi}{J}. \tag{4}$$

Consider a solution, even in both ξ and η, of the type

$$\chi/J=\phi_n(\eta)\cos n\xi. \tag{5}$$

The transformed biharmonic equation then reduces on substitution to an ordinary differential equation

$$\left\{\frac{d^4}{d\eta^4}-2(n^2+1)\frac{d^2}{d\eta^2}+(n^2-1)^2\right\}\phi_n(\eta)=0. \tag{6}$$

[1] R. C. J. Howland and R. C. Knight, "Stress functions for a plate containing groups of circular holes," Phil. Trans. **A 238**, 357 (1939).
[2] A. E. Green, "General biharmonic analysis for a plate containing circular holes," Proc. Roy. Soc. **A 176**, 121 (1940).
[3] G. B. Jeffery, "Plane stress and plane strain in bipolar coordinates," Phil. Trans. **A 221**, 265 (1921).

Consequently, the solution even in η is

$$\phi_n(\eta)=A_n\cosh(n+1)\eta+B_n\cosh(n-1)\eta, \quad (7)$$

where A_n and B_n are parametric coefficients. This solution holds true for any n except $n=0$. The solution in the latter case, however, will be rejected partially on the ground that it is redundant to stresses and partially that it gives rise to multi-valued displacements. There are also some even solutions of Eq. (4), which are not included in (7), for instance,

$$\chi/J=K(\cosh\eta-\cos\xi)\log(\cosh\eta-\cos\xi), \quad (8)$$

K being a parametric coefficient. It is found convenient to include this solution also in the solution concerned presently. A more general solution is, therefore,

$$\chi_1/J=K(\cosh\eta-\cos\xi)\log(\cosh\eta-\cos\xi)+\sum_{n=1}^{\infty}\phi_n(\eta)\cos n\xi. \quad (9)$$

The stresses derived from χ_1 are as follows:

$$\begin{aligned}a\widehat{\eta\eta}&=\left\{(\cosh\eta-\cos\xi)\frac{\partial^2}{\partial\xi^2}-\sinh\eta\frac{\partial}{\partial\eta}\right.\\&\quad\left.-\sin\xi\frac{\partial}{\partial\xi}+\cosh\eta\right\}\frac{\chi_1}{J}\\&=-\tfrac{1}{2}K(\cosh2\eta-2\cosh\eta\cos\xi-\cos2\xi)\\&\quad+\phi_1(\eta)+\tfrac{1}{2}\sum_{n=1}^{\infty}\{(n-1)(n-2)\phi_{n-1}(\eta)\\&\quad-2(n^2-1)\phi_n(\eta)\cosh\eta\\&\quad+(n+1)(n+2)\phi_{n+1}(\eta)\\&\quad-2\phi_n'(\eta)\sinh\eta\}\cos n\xi, \quad (10)\end{aligned}$$

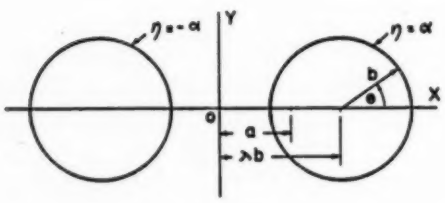


FIG. 1. Two equal holes.

$$\begin{aligned}a\widehat{\xi\xi}&=\left\{(\cosh\eta-\cos\xi)\frac{\partial^2}{\partial\eta^2}-\sinh\eta\frac{\partial}{\partial\eta}\right.\\&\quad\left.-\sin\xi\frac{\partial}{\partial\xi}+\cos\xi\right\}\frac{\chi_1}{J}\\&=\tfrac{1}{2}K(\cosh2\eta-2\cosh\eta\cos\xi+\cos2\xi)\\&\quad+\phi_1(\eta)-\tfrac{1}{2}\phi_1''(\eta)-\tfrac{1}{2}\sum_{n=1}^{\infty}\{\phi_{n-1}''(\eta)\\&\quad-2\phi_n''(\eta)\cosh\eta+\phi_{n+1}''(\eta)\\&\quad+(n-2)\phi_{n-1}(\eta)+2\phi_n'(\eta)\sinh\eta\\&\quad-(n+2)\phi_{n+1}(\eta)\}\cos n\xi, \quad (11)\end{aligned}$$

$$\begin{aligned}a\widehat{\xi\eta}&=-(\cosh\eta-\cos\xi)\frac{\partial^2}{\partial\xi\partial\eta}\left(\frac{\chi_1}{J}\right)\\&=-K\sinh\eta\sin\xi-\tfrac{1}{2}\sum_{n=1}^{\infty}\{(n-1)\phi_{n-1}'(\eta)\\&\quad-2n\phi_n'(\eta)\cosh\eta\\&\quad+(n+1)\phi_{n+1}'(\eta)\}\sin n\xi. \quad (12)\end{aligned}$$

Prime signs have the usual meaning of differentiation.

It is seen that all the stresses derived from χ_1 vanish at infinity, provided that $\chi_1/J=0$ when $\xi,\eta=0$. This condition leads to

$$\sum_{n=1}^{\infty}(A_n+B_n)=0. \quad (13)$$

When the holes are absent, the stress system in the (infinite) plate will be supposed to be specified by a symmetrical stress function χ_0. The method of solution when two equal holes are present is to add to χ_0 the even biharmonic function χ_1, as shown in (9), which satisfies the further condition (13) so as to give no stress at infinity. The parametric coefficients involved in the function χ_1 are adjusted to satisfy the boundary conditions at the edges of the holes. For convenience, a factor aT will be introduced to χ_1 to render these coefficients dimensionless. The required stress function is thus constructed in the form:

$$\chi=\chi_0+aT\chi_1. \quad (14)$$

Now, let the edges of the holes be denoted by $\eta=\pm\alpha$, respectively, such that their centers are located symmetrically on the x axis (see Fig. 1), and let the normal and the tangential stresses

derived from χ_0 at $\eta=\pm\alpha$ be denoted by

$$\widehat{\eta\eta}_0=T\sum_{n=0}^{\infty}c_n\cos n\xi,\quad \widehat{\xi\eta}_0=T\sum_{n=1}^{\infty}b_n\sin n\xi. \tag{15}$$

Consequently, the boundary condition of no traction at the edges of holes (i.e., $\widehat{\eta\eta}=0$, $\widehat{\xi\eta}=0$, when $\eta=\pm\alpha$) leads to

$$\left.\begin{aligned} -2c_0&=2\phi_1(\alpha)-K\cosh 2\alpha,\\ \text{and for } n\geqslant 1&\\ -2nc_n&=\psi_{n-1}-2\psi_n\cosh\alpha+\psi_{n+1}\\ &\quad -2\psi_n'\sinh\alpha+2K(\delta_{1n}\cosh\alpha-\delta_{2n}),\\ 2b_n&=\psi_{n-1}'-2\psi_n'\cosh\alpha\\ &\quad +\psi_{n+1}'+2K\delta_{1n}\sinh\alpha, \end{aligned}\right\}\tag{16}$$

where δ_{mn} is Kronecker delta which is equal to unity when $m=n$, or to zero when $m\neq n$, and

$$\psi_n=(n-1)n(n+1)\phi_n(\alpha),\quad \psi_n'=n\phi_n'(\alpha). \tag{17}$$

The systems of equations may be replaced by

$$\left.\begin{aligned} 2\phi_1(\alpha)&=K\cosh 2\alpha-2c_0,\\ \psi_1'&=2Ke^{-\alpha}\sinh\alpha-2\sum_{m=1}^{\infty}b_m e^{-m\alpha}, \end{aligned}\right\}\tag{18}$$

and for $n\geqslant 2$ by

$$\left.\begin{aligned} \psi_n'\sinh\alpha&=\psi_1'\sinh n\alpha\\ &\quad -2K\sinh(n-1)\alpha\sinh\alpha\\ &\quad +2\sum_{m=1}^{n-1}b_m\sinh(n-m)\alpha,\\ \psi_n\sinh\alpha&=\psi_1'(n\cosh n\alpha-\coth\alpha\sinh n\alpha)\\ &\quad +K\{(n+1)\sinh(n-2)\alpha\\ &\quad -(n-1)\sinh n\alpha\}\\ &\quad -2\sum_{m=1}^{n-1}\{(n-m)b_m\cosh(n-m)\alpha\\ &\quad -(mc_m+b_m\coth\alpha)\\ &\quad \times\sinh(n-m)\alpha\}. \end{aligned}\right\}\tag{19}$$

The coefficients A_n and B_n can thus be expressed explicitly in terms of c_n, b_n, and K. An additional equation is supplied by the condition (13) for the determination of K. The problem is thus solved in general for any even stress system acting in the plate.

TABLE I. The maximum stresses ($T=1$).

λ	(1) All-around tension		(2) Longitudinal tension	(3) Transverse tension	
	$\theta=0$	$\theta=\pi$	$\theta=\pm\pi/2$	$\theta=0$	$\theta=\pi$
1	2.894	∞	2.569	3.869	∞
1.5	2.255	2.887	2.623	3.151	3.264
2	2.158	2.411	2.703	3.066	3.020
3	2.080	2.155	2.825	3.020	2.992
5	2.033	2.049	2.927	3.004	2.997
8	2.014	2.018	2.970	3.001	2.999
∞	2.000	2.000	3.000	3.000	3.000

Three fundamental cases will be considered in the sequel, namely, (1) the plate is under an all-around tension T, so that we have

$$\chi_0=\tfrac{1}{2}T(x^2+y^2),$$

or

$$\chi_0/aTJ=\tfrac{1}{2}(\cosh\eta+\cos\xi); \tag{20}$$

(2) the plate is under a longitudinal tension T,

$$\chi_0=\tfrac{1}{2}Ty^2,$$

or

$$\chi_0/aTJ=\tfrac{1}{2}\sin^2\xi/(\cosh\eta-\cos\xi); \tag{21}$$

and (3) the plate is under a transverse tension T,

$$\chi_0=\tfrac{1}{2}Tx^2,$$

or

$$\chi_0/aTJ=\tfrac{1}{2}\sinh^2\eta/(\cosh\eta-\cos\xi). \tag{22}$$

The stresses at $\eta=\pm\alpha$ derived from χ_0 are then given by

$$\widehat{\eta\eta}/T=\widehat{\xi\xi}/T=1,\quad \widehat{\xi\eta}/T=0, \tag{23}$$

$$\left.\begin{aligned} \widehat{\eta\eta}/T&=(1-\cosh\alpha\cos\xi)^2/(\cosh\alpha-\cos\xi)^2,\\ \widehat{\xi\xi}/T&=\sinh^2\alpha\sin^2\xi/(\cosh\alpha-\cos\xi)^2,\\ \widehat{\xi\eta}/T&=-\sinh\alpha\sin\xi(1-\cosh\alpha\cos\xi)/\\ &\qquad (\cosh\alpha-\cos\xi)^2, \end{aligned}\right\}\tag{24}$$

and

$$\left.\begin{aligned} \widehat{\eta\eta}/T&=\sinh^2\alpha\sin^2\xi/(\cosh\alpha-\cos\xi)^2,\\ \widehat{\xi\xi}/T&=(1-\cosh\alpha\cos\xi)^2/(\cosh\alpha-\cos\xi)^2,\\ \widehat{\xi\eta}/T&=\sinh\alpha\sin\xi(1-\cosh\alpha\cos\xi)/\\ &\qquad (\cosh\alpha-\cos\xi)^2, \end{aligned}\right\}\tag{25}$$

respectively. Hence, we have in the first case

$$c_0=1,\quad c_n=b_n=0.\quad (n\geqslant 1) \tag{26}$$

In the other two cases, we expand the first and the third expressions, respectively, into Fourier series between $\xi=-\pi$ and $\xi=\pi$ with the aid of the following integrals, for $n\geqslant 0$,

$$\left.\begin{aligned}\int_{-\pi}^{\pi}\frac{\cos n\xi\, d\xi}{\cosh\alpha-\cos\xi}&=\frac{2\pi e^{-n\alpha}}{\sinh\alpha},\\ \int_{-\pi}^{\pi}\frac{\cos n\xi\, d\xi}{(\cosh\alpha-\cos\xi)^2}&=\frac{2\pi e^{-n\alpha}(n\sinh\alpha+\cosh\alpha)}{\sinh^3\alpha}.\end{aligned}\right\}(27)$$

Then we readily have

$$\left.\begin{aligned}c_0&=e^{-\alpha}\cosh\alpha,\\ c_n=b_n&=2e^{-n\alpha}\sinh\alpha(n\sinh\alpha-\cosh\alpha),\\ &\qquad(n\geqslant 1)\end{aligned}\right\}(28)$$

and

$$\left.\begin{aligned}c_0&=e^{-\alpha}\sinh\alpha,\\ c_n=b_n&=-2e^{-n\alpha}\sinh\alpha(n\sinh\alpha-\cosh\alpha),\\ &\qquad(n\geqslant 1)\end{aligned}\right\}(29)$$

respectively. With these coefficients we obtain

$$\left.\begin{aligned}2\phi_1(\alpha)&=K\cosh 2\alpha-2c_0,\\ \psi_n e^{n\alpha}&=-2K(n\sinh\alpha+\cosh\alpha)\\ &\qquad\pm n(n^2-1)\sinh\alpha,\quad(n\geqslant 2)\\ \psi_n' e^{n\alpha}&=2K\sinh\alpha\mp n(n\sinh\alpha-\cosh\alpha).\\ &\qquad(n\geqslant 1)\end{aligned}\right\}(30)$$

Wherever the sign is ambiguous, the upper one goes with the longitudinal tension case and the lower with the transverse tension case. In the all-around tension case, the terms with ambiguous signs are all absent. Hence, we have for $n\geqslant 1$,

$$\left.\begin{aligned}A_n&=\frac{2K(e^{-n\alpha}\sinh n\alpha+ne^{-\alpha}\sinh\alpha)\pm n(n+1)(e^{-n\alpha}\cosh n\alpha-ne^{-\alpha}\sinh\alpha)}{n(n+1)(\sinh 2n\alpha+n\sinh 2\alpha)}\\ B_n&=-\frac{2K(e^{-n\alpha}\sinh n\alpha+ne^{\alpha}\sinh\alpha)\pm n(n-1)(e^{-n\alpha}\cosh n\alpha-ne^{\alpha}\sinh\alpha)}{n(n-1)(\sinh 2n\alpha+n\sinh 2\alpha)},\end{aligned}\right\}(31)$$

with the exception that

$$B_1=\tfrac{1}{2}(K\tanh\alpha\cosh 2\alpha-2c_0\mp e^{-2\alpha}\coth 2\alpha).$$

Moreover, in the all-around tension case, K is determined from

$$K\left\{\tfrac{1}{2}+\tanh\alpha\sinh^2\alpha-4\sum_{n=2}^{\infty}\frac{e^{-n\alpha}\sinh n\alpha+n\sinh\alpha(n\sinh\alpha+\cosh\alpha)}{n(n^2-1)(\sinh 2n\alpha+n\sinh 2\alpha)}\right\}=1. \tag{32}$$

But in the longitudinal and the transverse tension cases, unity on the right-hand side of Eq. (32) is to be replaced by the quantities

$$\tfrac{1}{2}\pm\tfrac{1}{2}\mp 2\sinh^2\alpha\sum_{n=1}^{\infty}n/(\sinh 2n\alpha+n\sinh 2\alpha). \tag{33}$$

Most important is the stress at the edges of holes. With the same distinction to ambiguous signs as before, this stress is found to be given by

$$\frac{\widehat{\xi\xi}_a}{T}=2(\cosh\alpha-\cos\xi)\left\{K\sinh\alpha\left(1+4\sum_{n=1}^{\infty}\frac{\sinh n\alpha\cos n\xi}{\sinh 2n\alpha+n\sinh 2\alpha}\right)\right.$$
$$\left.\mp 2\sum_{n=1}^{\infty}\frac{n(n\sinh n\alpha\sinh\alpha-\cosh n\alpha\cosh\alpha)\cos n\xi}{\sinh 2n\alpha+n\sinh 2\alpha}\right\}. \tag{34}$$

For convenience, we shall define η and θ as shown in Fig. 1, which are connected to (ξ, α) by

$$\cosh\alpha=\lambda, \quad \cos\xi=(1+\lambda\cos\theta)/(\lambda+\cos\theta). \quad (35)$$

It appears that there are two points of maximum (tensile) stress at the edge of each hole. In the all-around and the transverse tension cases, they are at $\theta=0$ (or $\xi=0$) and at $\theta=\pi$ (or $\xi=\pi$). In the longitudinal tension case, these points are very nearly at $\theta=\pm\pi/2$, or more precisely at $|\theta|$ slightly less than $\pi/2$. They shift towards $\theta=\pm\pi/2$ when λ increases. For simplicity in calculation, the stress at $\theta=\pm\pi/2$ in the latter case may be taken as the maximum stress without any appreciable error.

The limiting case $\alpha=\infty$ (or $\lambda=\infty$) corresponds to a single hole in the plate. The results in this case are well-known. On the other hand, in the limiting case $\alpha=0$ (or $\lambda=1$), the two holes in the plate become tangential to each other. It can be shown without much difficulty that when $\alpha=0$, the series in Eq. (32) takes the following limiting value

$$2\sum_{n=2}^{\infty}\frac{e^{-n\alpha}\sinh n\alpha+n\sinh\alpha(n\sinh\alpha+\cosh\alpha)}{n(n^2-1)(\sinh 2n\alpha+n\sinh 2\alpha)}$$

$$=\sum_{n=2}^{\infty}\frac{1}{n(n^2-1)}-2\alpha^2\int_0^{\infty}\frac{(\sinh^2\phi-\phi^2)d\phi}{\phi^3(\sinh 2\phi+2\phi)}.$$

Since the sum of the series on the right-hand side of the expression is equal to $\frac{1}{4}$, Eq. (32) thus becomes in the limit

$$4(\alpha^2K)\int_0^{\infty}\frac{(\sinh^2\phi-\phi^2)d\phi}{\phi^3(\sinh 2\phi+2\phi)}=1. \quad (36)$$

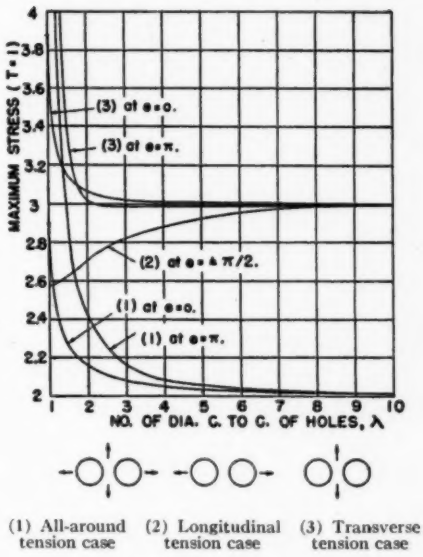
Also, the quantities in (33) become in the limit

$$\tfrac{1}{2}\pm\tfrac{1}{2}\mp 2\int_0^{\infty}\frac{\phi\,d\phi}{\sinh 2\phi+2\phi}. \quad (37)$$

The maximum stresses take the following limiting values when $\alpha=0$: (1) in the all-around tension case at $\xi=0$,

$$\frac{\widehat{\xi\xi}_0}{T}=4(\alpha^2K)\int_0^{\infty}\frac{\sinh\phi\,d\phi}{\sinh 2\phi+2\phi}, \quad (38)$$

(2) in the longitudinal tension case at $\xi=\alpha$ (or $\theta=\pm\pi/2$),



(1) All-around tension case (2) Longitudinal tension case (3) Transverse tension case

FIG. 2. The maximum stress.

$$\frac{\widehat{\xi\xi}_0}{T}=8(\alpha^2K)\int_0^{\infty}\frac{\sinh\phi\cos\phi\,d\phi}{\sinh 2\phi+2\phi}$$

$$+4\int_0^{\infty}\frac{\phi(\cosh\phi-\phi\sinh\phi)\cos\phi\,d\phi}{\sinh 2\phi+2\phi}, \quad (39)$$

and (3) in the transverse tension case at $\xi=0$,

$$\frac{\widehat{\xi\xi}_0}{T}=4(\alpha^2K)\int_0^{\infty}\frac{\sinh\phi\,d\phi}{\sinh 2\phi+2\phi}$$

$$+2\int_0^{\infty}\frac{\phi(\phi\sinh\phi-\cosh\phi)d\phi}{\sinh 2\phi+2\phi}. \quad (40)$$

Values of the maximum stresses are calculated and shown, for $T=1$, in the accompanying table. Figure 2 shows the results graphically.

It may be noted that the integrals in the limiting case $\alpha=0$ can be conveniently evaluated with the aid of the numerical values of the following integrals, for $n\geq 1$,

$$I_n=\frac{2^n}{n!}\int_0^{\infty}\frac{\phi^n d\phi}{\sinh 2\phi+2\phi}, \quad I_{\infty}=1, \quad (41)$$

which have been calculated by Howland.[4] For examples:

[4] R. C. J. Howland, "On the stresses in the neighborhood of a circular hole in a strip under tension," Phil. Trans. **A 229**, 67 (1930), Table I.

$$\left.\begin{aligned}
\int_0^\infty \frac{\sinh\phi\, d\phi}{\sinh 2\phi + 2\phi} &= \sum_{n=0}^{\infty} \frac{I_{2n+1}}{2^{2n+1}} = 0.52686, \\
\int_0^\infty \frac{\sinh\phi\cos\phi\, d\phi}{\sinh 2\phi + 2\phi} &= \sum_{n=0}^{\infty} (-1)^{n(n+1)/2} \frac{I_{2n+1}}{2^{n+1}} = 0.13474, \\
\int_0^\infty \frac{(\sinh^2\phi - \phi^2)d\phi}{\phi^3(\sinh 2\phi + 2\phi)} &= \sum_{n=1}^{\infty} \frac{I_{2n-1}}{n(n+1)(2n+1)} = 0.18203.
\end{aligned}\right\} \quad (42)$$

An Alignment Chart Giving the Polarization Correction of Equi-Inclination Weissenberg Photographic Intensities

W. L. Bond
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received August 15, 1947)

A nomograph is presented which gives, in terms of film coordinates μ and Υ, the polarization factors to be applied to intensities taken from equi-inclination Weissenberg photographs to correct them for polarization. The remaining correction, the Lorentz factor, merely involves division by $\cos^2\mu \sin\Upsilon$.

In determining the structure factors F or F^2 from observed intensities of spots on equi-inclination Weissenberg photographs,* a correction for polarization is necessary. The computation of the polarization correction factor, $1/p$, is somewhat tedious. Here, $F^2 \sim 1/L \cdot 1/p \cdot I_0$, where I_0 is the spot intensity

$$p = \frac{1+\cos^2 2\theta}{2}$$

and

$$L \text{ (the Lorentz factor)} = \frac{1}{\cos^2\mu \sin\gamma}.$$

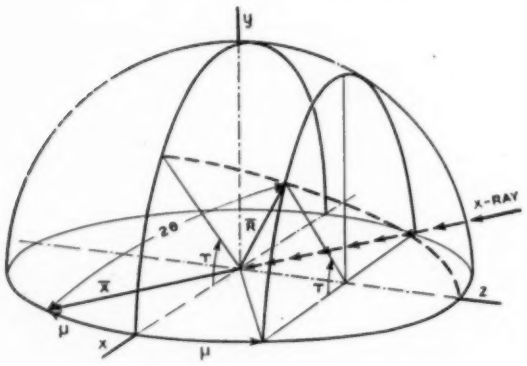


Fig. 1.

Knowing μ the inclination angle and Υ the off-axis angle of the reflection, one can compute 2θ, the ray deviation, and hence $1/p$. This can be done very conveniently and with sufficient accuracy with an alignment chart. This paper presents such a chart which gives the polarization correction for equi-inclination Weissenberg photographs.

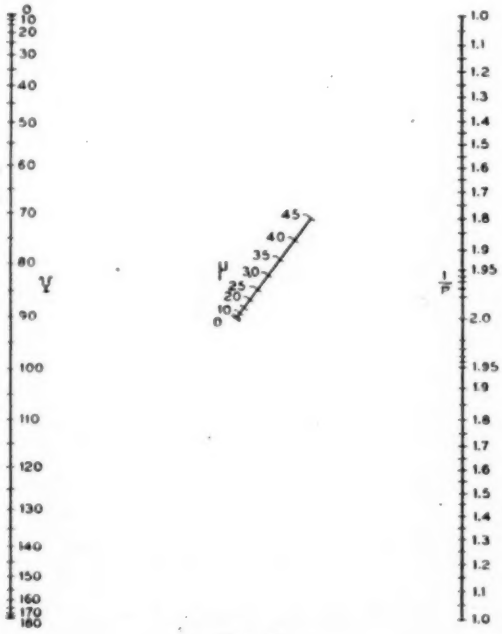


Fig. 2.

* M. J. Buerger and G. E. Klein, "Correction of x-ray diffraction intensities for Lorentz and polarization factors," J. App. Phys. **16**, 408 (1945).

A study of Fig. 1 reveals that μ, Υ, and θ are related by means of the equation:

$$\cos 2\theta = \cos^2\mu \cos\Upsilon - \sin^2\mu$$

(see Appendix). Writing this as a determinant we have

$$\begin{vmatrix} \cos^2\mu \cos\Upsilon & 1 & 0 \\ \sin^2\mu & 1 & 1 \\ -\cos 2\theta & 0 & 1 \end{vmatrix} = 0,$$

which easily reduces to the standard form:

$$\begin{vmatrix} \dfrac{\cos^2\mu}{1+\cos^2\mu} & \dfrac{\sin^2\mu}{1+\cos^2\mu} & 1 \\ 1 & \cos\Upsilon & 1 \\ 0 & -\cos 2\theta & 1 \end{vmatrix} = 0.$$

Since p and θ are related through

$$\cos 2\theta = (2p-1)^{\frac{1}{2}},$$

we can mark the θ curve in terms of the correction factor $1/p$ instead of θ. This gives us the alignment chart, Fig. 2.

APPENDIX

The incoming x-ray travels along the unit vector $\bar{X}$, where, on rectangular unit axes x, y, and z, the components are

$$\bar{X} = \begin{Vmatrix} \cos\mu \\ 0 \\ -\sin\mu \end{Vmatrix}.$$

The components of $\bar{R}$, the unit vector along which the reflected ray travels, are readily seen to be

$$\bar{R} = \begin{Vmatrix} \cos\mu \cos\Upsilon \\ \cos\mu \sin\Upsilon \\ \sin\mu \end{Vmatrix}.$$

(Consider $\bar{R}$ as given in terms of spherical coordinate angles $(90-\mu)$ and Υ.) As $\bar{R}\cdot\bar{X}$, the scalar product of $\bar{R}$ and $\bar{X}$, is $\cos 2\theta$, we have

$$\cos 2\theta = \cos^2\mu \cos\Upsilon - \sin^2\mu.$$

A Simple Procedure for the Making of Alignment Charts

W. L. Bond
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received August 15, 1947)

Simplified directions are given for the making of alignment charts from equations in three variables. A fixed determinant is used to put the equation in determinantal form. The determinant is manipulated to put it in standard line form; three standard types of distortion factors are used to produce a chart of convenient shape.

THIS paper presents the procedure that was used to prepare the alignment chart or nomograph given in the preceding paper. The methods given here suffice to prepare alignment charts for many equations encountered in physics and engineering.

By similar triangles, Fig. 1, we see that if $A-B$ is a straight line, then:

$$\frac{x_1-x_2}{y_1-y_2} = \frac{x_1-x_3}{y_1-y_3}.$$

This expands to:

$$x_1(y_2-y_3)+x_2(y_3-y_1)+x_3(y_1-y_2)=0. \quad (1)$$

A useful identity* is:

$$a+b+c = \begin{vmatrix} a & 1 & 0 \\ -b & 1 & 1 \\ c & 0 & 1 \end{vmatrix} \cdots. \quad (2)$$

Application of Eq. (2) to Eq. (1) yields:

$$\begin{vmatrix} x_1(y_2-y_3) & 1 & 0 \\ x_2(y_1-y_3) & 1 & 1 \\ x_3(y_1-y_2) & 0 & 1 \end{vmatrix} = 0. \quad (3)$$

To illustrate the procedure we now reduce Eq. (3) to the standard determinant equation of a line by operating on the determinant, with operations that do not change the value of the determinant.

* Expansion of the determinant:

$$\Delta = \begin{vmatrix} a & 1 & 0 & 0 & 0 \\ -b & 1 & 1 & 0 & 0 \\ c & 0 & 1 & 1 & 0 \\ -d & 0 & 0 & 1 & 1 \\ e & 0 & 0 & 0 & 1 \end{vmatrix}$$

gives $\Delta = a+b+c+d+e$. Equation (1) is a special case of this identity. The identity can be extended indefinitely as can be seen by considering the bottom row minors of the determinant that results by adding another term. The value of the new determinant is obviously unity times the old determinant plus the new term, (f in this case) times a unit determinant.

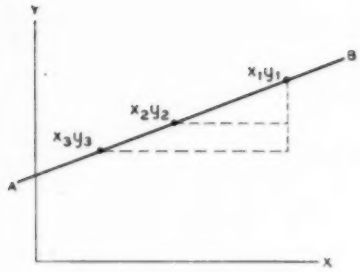


FIG. 1.

The chief operation is that of adding any row (or column) multiplied by a common factor, to any other row (or column). Since the determinant is equal to zero, we may also simply multiply or divide any row (or column) by a common factor. If we divide the top row by y_2-y_3 then multiply the second column by y_2-y_3, the result is merely to shift this factor, giving:

$$\begin{vmatrix} x_1 & 1 & 0 \\ x_2(y_1-y_3) & y_2-y_3 & 1 \\ x_3(y_1-y_2) & 0 & 1 \end{vmatrix} = 0.$$

Similarly, operating on the bottom row and last column then dividing the second row by (y_1-y_3), we obtain:

$$\begin{vmatrix} x_1 & 1 & 0 \\ x_2 & \dfrac{y_2-y_3}{y_1-y_3} & \dfrac{y_1-y_2}{y_1-y_3} \\ x_3 & 0 & 1 \end{vmatrix} = 0.$$

Adding the second column to the third column then multiplying the second column by (y_1-y_3):

$$\begin{vmatrix} x_1 & y_1-y_3 & 1 \\ x_2 & y_2-y_3 & 1 \\ x_3 & 0 & 1 \end{vmatrix} = 0.$$

Finally, adding the last column multiplied by y_3 to the second column we have the standard form:

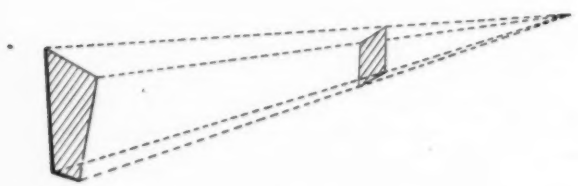


FIG. 2.

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0. \tag{4}$$

Let us assume that x_1 and y_1 are functions of a parameter u, and, further, x_2 and y_2 are functions of a parameter v, and, finally, x_3 and y_3 are functions of a parameter w, as:

$$\begin{aligned} x_1 &= x_1(u), \quad y_1 = y_1(u), \\ x_2 &= x_2(v), \quad y_2 = y_2(v), \\ x_3 &= x_3(w), \quad y_3 = y_3(w). \end{aligned} \tag{5}$$

If now u, v, and w are not independent but are related by an equation such as

$$f(u,\ v,\ w) = 0, \tag{6}$$

then if Eq. (6) can be expressed in the form of Eq. (4), corresponding values of u, v, and w will lie on a straight line.

As an example, consider the relation

$$\cos 2\theta = \cos^2\mu \cos\Upsilon - \sin^2\mu, \tag{7}$$

where θ varies from 0 to 90°, Υ varies from 0 to 180°, and μ varies from 0 to 45°. By use of Eq. (2), Eq. (7) becomes:

$$\begin{vmatrix} \cos^2\mu\cos\Upsilon & 1 & 0 \\ \sin^2\mu & 1 & 1 \\ -\cos 2\theta & 0 & 1 \end{vmatrix} = 0 \tag{8}$$

$$= \begin{vmatrix} \cos^2\mu & 1 & 0 \\ \sin^2\mu & \cos\Upsilon & \cos 2\theta \\ -1 & 0 & 1 \end{vmatrix} = 0$$

$$= \begin{vmatrix} \cos^2\mu & 1 & 0 \\ \sin^2\mu & \cos\Upsilon & \cos 2\theta \\ -1-\cos^2\mu & -1 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{\cos^2\mu}{1+\cos^2\mu} & 1 & 0 \\ \dfrac{\sin^2\mu}{1+\cos^2\mu} & \cos\Upsilon & -\cos 2\theta \\ 1 & 1 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} \dfrac{\cos^2\mu}{1+\cos^2\mu} & \dfrac{\sin^2\mu}{1+\cos^2\mu} & 1 \\ 1 & \cos\Upsilon & 1 \\ 0 & -\cos 2\theta & 1 \end{vmatrix}. \tag{9}$$

There are many ways of putting (8) in standard form. If, in the determinant following (8) the top row had been added to the bottom row instead of subtracted, the resulting terms

$$\frac{\cos^2\mu}{1-\cos^2\mu} \quad \text{and} \quad \frac{\sin^2\mu}{1-\cos^2\mu}$$

would become infinite for $\mu=0°$. Since μ must be charted for all values between 0° and 45°, we use the subtraction result instead of the addition result.

Using cross section paper we now draw the curve

$$x=\frac{\cos^2\mu}{1+\cos^2\mu}, \quad y=\frac{\sin^2\mu}{1+\cos^2\mu},$$

marking the values of μ along this curve. We similarly draw the curves $x=1$, $y=\cos\Upsilon$ and $x=0$, $y=-\cos 2\theta$. These three curves constitute the alignment chart.

Commonly the shape of the desired part of the curves is awkward. It is generally desired that the area to be used be nearly square. This can be accomplished through distortions that leave all straight lines straight, such as given by the projection, Fig. 2.

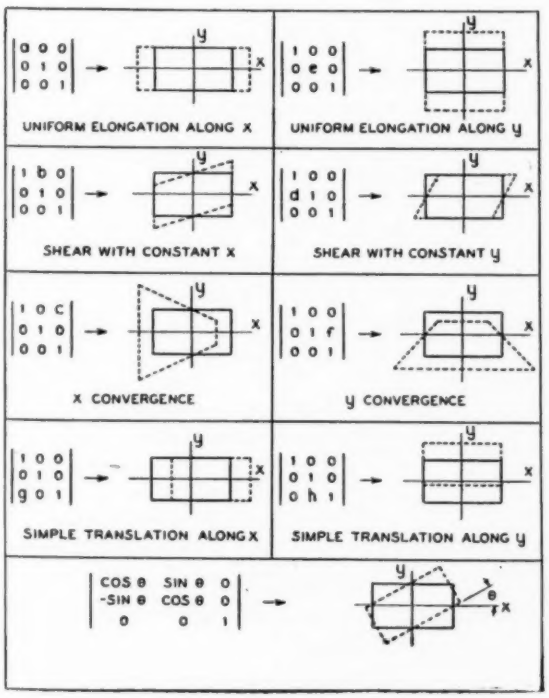


FIG. 3.

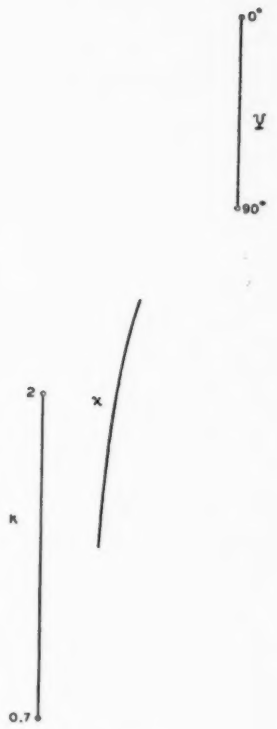


FIG. 4.

This general distortion can be achieved by multiplying the determinant in the form of Eq. (8) by a determinant made up of 9 constants. These constants can later be evaluated to give the final chart the shape desired. It is generally easier to perform the distortion step by step, operating on the standard form, Eq. (9), with a postfactor determinant.

Since

$$\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \cdot \begin{vmatrix} a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} ax_1 & y_1 & 1 \\ ax_2 & y_2 & 1 \\ ax_3 & y_3 & 1 \end{vmatrix}.$$

We see that the result of this post multiplier is a uniform elongation along x, as in Fig. 3, upper left. As a guide to the step by step distortions we give in Fig. 3 a few simple operator determinants that can be combined to produce any distortion desired. (Negative values of the constants reverse the sense of distortion.)

As an example, let us prepare a nomograph for the equation:

$$\cos^2\Upsilon\cos^2\chi+1-\frac{2}{k}\cos\chi=0,$$

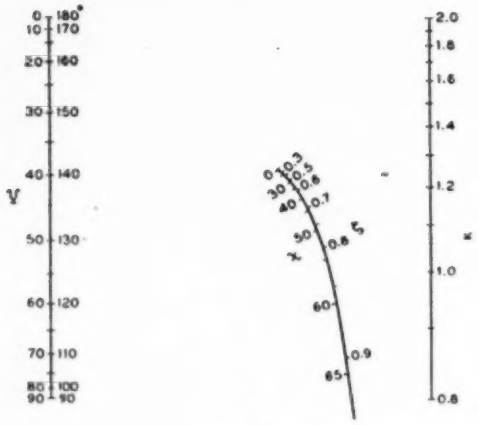


FIG. 5.

where ϒ runs from 0 to 180° and χ from 0 to 65°. This is easily given the standard form:

$$\begin{vmatrix} \dfrac{-1}{1+\cos\chi} & \dfrac{-1}{(1+\cos\chi)\cos\chi} & 1 \\ 0 & \cos^2\Upsilon & 1 \\ -1 & -\dfrac{2}{k} & 1 \end{vmatrix} = 0. \quad (10)$$

This determinant gives the awkward plot of Fig. 4. It is decided to square up the portion $0.8 \leq k \leq 2$, $0 \leq \gamma \leq 90$. The distortions of Fig. 3 may be applied in any order, but sometimes one order is easier than other orders, hence we first make the figure into a parallelogram, then shear it into a rectangle, and finally equalize the length and width of the diagram. We first apply the x convergence distortion to make this figure a parallelogram. Determinant (10) multiplied by the postfactor

$$\begin{vmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

gives

$$\begin{vmatrix} \dfrac{-1}{1+\cos\chi} & \dfrac{-1}{(1+\cos\chi)\cos\chi} & \dfrac{-c}{1+\cos\chi}+1 \\ 0 & \cos^2\Upsilon & 1 \\ -1 & -\dfrac{2}{k} & -c+1 \end{vmatrix}. \quad (11)$$

We must now divide each row by its last term to re-establish standard form. Fortunately the ϒ term still occupies a unit length along the y axis because ϒ values are unchanged (see Fig. 3 and note that $x=0$ for all values of ϒ). The y axis length of the k curve is found to be

$$-\left(\frac{2}{2(c+1)}-\frac{2}{0.8(c+1)}\right),$$

which is unity if $c=-1/2$.

We now substitute $c=-\frac{1}{2}$ into determinant (11) and apply the distorting post factor

$$\begin{vmatrix} 1 & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

to make the figure a rectangle:

$$\begin{vmatrix} \dfrac{-1}{1.5+\cos\chi} & \dfrac{-b\cos\chi-1}{(1.5+\cos\chi)\cos\chi} & 1 \\ 0 & \cos^2\Upsilon & 1 \\ -2/3 & \dfrac{-2b}{3}-\dfrac{4}{3k} & 1 \end{vmatrix} = 0.$$

For $k=2$ to have the same y value as has ϒ = 0 we must have $b=-5/2$, whence we get, upon multiplying the first column by $-\frac{3}{2}$ (to make the figure square):

$$\begin{vmatrix} 1.5 & 2.5\cos\chi-1 & 1 \\ \dfrac{1.5+\cos\chi}{0} & \dfrac{(1.5+\cos\chi)\cos\chi}{\cos^2\Upsilon} & 1 \\ 1 & \dfrac{1}{3}(5-4/k) & 1 \end{vmatrix} = 0.$$

This gives the chart of Fig. 5. The k of this chart when multiplied by sinϒ is the complete intensity correction factor for rotation x-ray photographs recorded on cylindrical film.

Characteristics of Helical Antennas Radiating in the Axial Mode

JOHN D. KRAUS AND J. CLAUDE WILLIAMSON
Communications Laboratory, Department of Electrical Engineering, Ohio State University, Columbus, Ohio
(Received July 22, 1947)

A helix is a fundamental form of antenna with many radiation modes. A recently reported mode, called an axial or beam mode, occurs for a relatively wide range of helix dimensions, in the region of 0.2 to 0.5 wave-lengths diameter and as high as 0.5 wave-lengths spacing between turns. The radiation is maximum in the direction of the helix axis and is nearly circularly polarized. This mode may persist with a given helix over a considerable frequency range. In this range the phase velocity of wave propagation along the helical conductor is reduced. An approximate expression for the field pattern of a single turn helix is developed. The pattern of a helix of a number of turns is then calculated as an array of such turns. Measured and calculated patterns show good agreement.

INDEX OF FREQUENTLY USED SYMBOLS

$A.R.$ = axial ratio of polarization ellipse
b = length of short conductor
c = velocity of light
D = diameter of helix (center-to-center)
D_λ = helix diameter in free space wave-lengths
E_ϕ = ϕ-component of E-vector
E_θ = θ-component of E-vector
g = side of square helix
L = helical length of 1 turn
n = number of turns
p = phase velocity factor = v/c
S = spacing between turns (center-to-center)
S_λ = spacing between turns in free space wave-lengths
SWR = standing wave ratio
v = phase velocity along helical conductor
α = pitch angle of helix
β = see Fig. 15B
ϵ = dielectric constant of medium (free space in present case)
λ = wave-length in free-space
μ = permeability of medium (free space in present case)
τ = tilt angle between radiation maximum and normal to conductor
$\omega = 2\pi$ (frequency)

INTRODUCTION

A CONDUCTOR wound as a helix or coil is a familiar form of circuit element. Its properties are well known at frequencies for which the circumference of the helix cylinder is small compared to the wave-length.*[1,2] A helical coil of such dimensions has also found application in the traveling wave tube.[3–5]

Recently it has been shown that when the circumference of the helix cylinder is increased to the order of one wave-length, the helix can be excited as a beam antenna.[6] The maximum radiation is in the direction of the helix axis and is nearly circularly polarized. This radiation mode will be referred to as an axial or beam mode.

Referring to Fig. 1, the diameter of a helix of circular cross section will be designated as D, the spacing as S, the helical length of one turn as L, and the pitch angle as α where

$$\alpha = \arctan(S/\pi D).$$

The relations between S, D, and α are shown in a particularly convenient way by a D–S chart as in Fig. 2. If D of a helix is fixed while S is reduced to zero, α becomes zero and the helix collapses to a loop. On the other hand, if S is fixed while D is reduced to zero, α becomes 90° and the helix straightens out into a linear conductor. Hence, the helix can be regarded as a general form of antenna of which loops and linear conductors are special cases. Loop and linear conductor antennas, as represented by the D and S axes of Fig. 2, have been extensively studied but the area between has not, except close to the origin.

* A complete bibliography for this case is too extensive to present here. References 1 and 2 are, however, representative.

[1] J. W. Nicholson, "The effective resistance and inductance of a helical coil," Phil. Mag. **19**, 77 (1910).

[2] Reinhold Rüdenberg, "Electromagnetic waves in transformer coils treated by Maxwell's equations," J. App. Phys. **12**, 219 (1941).

[3] Rudolf Kompfner, "The traveling wave tube as amplifier at microwaves," Proc. I.R.E. **35**, 124 (1947).

[4] J. R. Pierce, "Theory of the traveling wave tube", Proc. I.R.E. **35**, 111 (1947).

[5] L. J. Chu and D. Jackson, "Field theory of traveling wave tubes," Research Laboratory of Electronics, M.I.T., Tech. Report 38 (1947).

[6] John D. Kraus, "Helical beam antenna," Electronics **20**, No. 4, 109 (1947).

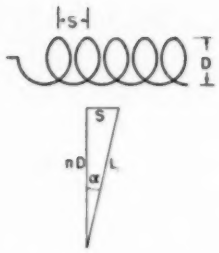


FIG. 1. Relation of helix dimensions.

This paper discusses some of the radiation modes occurring in the region between the D and S axes shown in Fig. 2 and in particular the axial mode. This mode can be excited over a relatively large region of helix dimensions. The region is defined approximately near its upper limit by the curve

$$D_\lambda = (2S_\lambda + 1)^{\frac{1}{2}}/\pi \qquad (1)$$

which is indicated in Fig. 2 as curve A. Outside the axial mode region a helix can be excited in many other radiation modes. For example, helices with dimensions which place it close to the axial mode region may exhibit conical patterns. Another mode has been described by Wheeler[7] in which radiation is maximum normal to the helix axis and is circularly polarized. To obtain this mode the E-vector resulting from currents on helix components parallel to the axis and the E-vector resulting from currents on helix components normal to the axis must be equal and in phase quadrature. If the phase velocity is infinite, D is small and the current is uniform along the helix, this condition is obtained

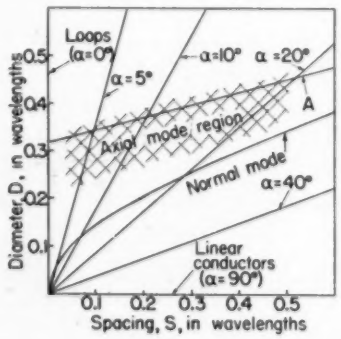


FIG. 2. Diameter-spacing chart for helix showing dimensions for various radiation modes.

[7] H. A. Wheeler, "Helical antenna for circular polarization," Proc. I.R.E., **35**, 1484 (1947).

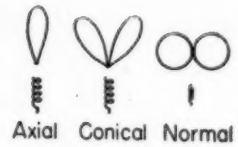


FIG. 3. Idealized patterns of helices radiating in axial, conical and normal modes.

approximately when

$$D_\lambda = (2S_\lambda)^{\frac{1}{2}}/\pi.^{**} \qquad (2)$$

The "normal mode" curve in Fig. 2 is a plot of this equation. This curve is only of practical significance near the origin for a simple helix,*** since the combination of a uniform current distribution and an infinite phase velocity is only readily approximated on simple helices of very short length. Idealized patterns of helices radiating in axial, conical, and normal modes are compared in Fig. 3.

CHARACTERISTICS OF THE AXIAL MODE

A helix may be readily excited in the axial mode by a number of arrangements. A very simple method is to connect one end of the helix to the inner conductor of a coaxial transmission line and terminate the outer conductor of the line in a ground plane. The ground plane may be flat or conical as suggested in Fig. 4. In the beam mode the axial component of the E-vector is substantially zero at the helix axis so that a conductor may be placed at the axis with little change in characteristics. Two arrangements with an axial conductor are shown in Fig. 4. Two helices wound in opposite directions (i.e., one left-handed and the other right-handed) are connected in series and placed side-by-side to form a "folded" helix as illustrated. A variation is the "double layer" helix in which two helices of unequal diameter and opposite screw direction are placed with axes coincident. A helix energized by a balanced or two-wire line is also shown in Fig. 4. The following discussion applies specifically to uniform helices energized by a coaxial line with the outer conductor terminated in a ground plane as in the top-left sketch of Fig. 4.

** This equation is equivalent to $A = pl$ as given by Wheeler where A = area of helix cross section, p = spacing, and $l = \lambda/2\pi$.

*** By simple helix is meant one of simple geometric form (as in Fig. 1), unmodified by auxiliary devices such as capacitors, phase changers, etc.

The experimental arrangement used is shown by the photograph of Fig. 5.

Measured Radiation Patterns

It is convenient to describe the radiation characteristics of a helix by two components of the E-vector, E_ϕ and E_θ. Their relation to the helix is illustrated in Fig. 6.

The helix in Fig. 5 has its axis horizontal and can be rotated in azimuth (ϕ). A linearly polarized transmitting array situated at a distance provides a substantially uniform field at the helix. When the helix is rotated in azimuth its ϕ-pattern is obtained in relative field intensity with the aid of a calibrated receiver. By reciprocity, this pattern is the same as with the helix transmitting. With the linear transmitting array horizontal, E_ϕ is measured and with the transmitting array vertical, E_θ is measured. Referring to Fig. 6, E_ϕ and E_θ in the following discussion are measured in the x–z plane as a function of ϕ.

Measured axial mode patterns are shown in Fig. 7 at several frequencies for the helix in Fig. 5. This helix consists of 7 turns of 1.27-cm o.d. copper tube, with $\alpha = 12°$, $D = 22.5$ cm, $S = 15$ cm, and an axial length from the ground plane of 112 cm. The ground plane used in all measurements is 66 cm in diameter. The patterns for both E_ϕ and E_θ are adjusted to the same maximum value. In general, E_ϕ and E_θ are of unequal maximum value, depending on the axial ratio of the polarization ellipse (see Fig. 12) and also on the orientation of the ellipse with respect to the reference coordinates at the frequency measured. Above about 520 Mc, the pattern exhibits a marked change, breaking up into many lobes. A typical pattern is illustrated in the upper half of Fig. 8. This pattern was measured at 630 Mc with a helix identical to the one in Fig. 5 except that it consists of 8 instead of 7 turns. When the frequency is too high or too low for the axial mode, the pattern is sometimes conical as illustrated in the lower half of Fig. 8. This pattern (E_θ) was measured with a 24° helix ($D = 0.40$ and S = 0.55 wave-lengths) of 4 turns at 556 Mc. The above measurements indicate that the axial mode of radiation characterized by a major axial lobe and small minor lobes persists over a frequency range of nearly 2 to 1

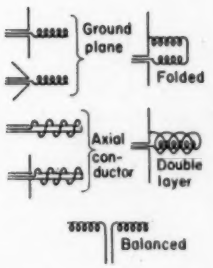


FIG. 4. Arrangements for energizing helix in axial mode.

for the helix in Fig. 5, with upper and lower limits in frequency at which the pattern becomes many lobed. A comparison between helices of 1.27-cm and 1-mm o.d. conductors indicates a greater frequency range of the axial mode for the thicker conductor.

A systematic set of measurements was made to determine the approximate range of diameters and spacings for which the axial mode is observed for helices of constant physical length. Four helices of about the same diameter ($D = 22.5$ cm) and length from ground plane to open end (123 cm) but of different pitch angles ($\alpha = 6°$, 12°, 18°, and 24°) were measured over a frequency range. They consisted of 15, 8, 5, and 3.9 turns respectively and were wound of 1.27 cm o.d. copper tube. With variable frequency, the dimensions in wave-lengths for a given helix move along the pitch-angle, α, line on a D–S chart.

The results of about 40 E_ϕ pattern measurements on the four helices are summarized in Fig. 9. The electrical length of the helices varied with frequency between about 1 and 2.5 wave-lengths as indicated by the dashed horizontal lines. The contours give the approximate variation of the beam-width of the patterns. The width is taken between half-power points (0.707

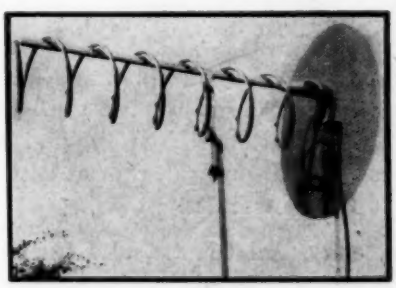


FIG. 5. 12° helix of 7 turns mounted on ground plane for pattern measurements. The helix diameter is 22.5 cm.

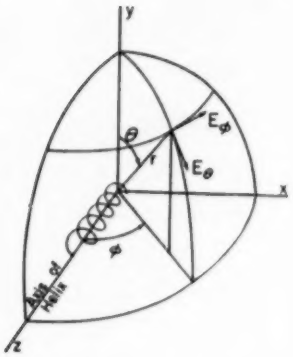


FIG. 6. Polar coordinates for field components.

points of maximum field intensity) regardless of whether these occur on the major lobe or on minor lobes. Thus, if the minor lobe level exceeds 71 percent of the major lobe maximum, the minor lobes are included in the beam-width. This definition is arbitrary but is convenient in this case to take into account a splitting up of the pattern into many lobes of nearly equal amplitude. In Fig. 9, contours of 30°, 40°, and 50° beam-widths are indicated. The region in which the side lobe level is a minimum is indicated by the cross-hatched area. In general, outside the 50° contour, the pattern has disintegrated into a conical or multi-lobed form.

The results of about 40 E_θ pattern measurements on the four helices are summarized in Fig. 10. Half-power beam-width contours of 35°, 40°, 50°, and 60°, are shown and the region of minimum side-lobe level is indicated by the cross-hatched area. Comparing Figs. 9 and 10, it is apparent that the E_ϕ patterns are in general sharper than the E_θ patterns.

The effect of helix length was studied by measuring patterns at a fixed frequency as a function of the number of turns. Results of such measurements are shown in Fig. 11 for a 12° helix of 1 to 10 turns ($D=23.3$ cm, $S=15.7$ cm). E_ϕ patterns at 400 Mc for 10, 4, and 1 turns are illustrated in Fig. 11, as are also the half-power beam widths for E_ϕ and E_θ at 400 and 500 Mc. In general, the beam-width decreases as the number of turns increases.

Phase Velocity, Axial Ratio, and Current Distribution

By measuring the distance between the maxima or minima of the standing waves of current on the helix, the phase velocity of wave propagation along the helix can be determined. The ratio of the phase velocity, v, along the helical conductor to the velocity of light, c, will be called $p(p=v/c)$. Measured values of p for the helix of Fig. 5 are presented in Fig. 12 for frequencies between 200 and 500 Mc. Below about 300 Mc p is approximately unity, but above 300 Mc it decreases to about 0.65 and then increases in an approximately straight-line fashion.

The ratio of the major to the minor axes of the polarization ellipse of the electric vector may be conveniently referred to as the axial ratio ($A.R.$). It is unity for circular polarization and infinite for linear (plane) polarization. To observe the axial ratio, the linearly polarized transmitting array, as described for the pattern measurements, may be rotated through all polarization angles; or for measurements only in the direction of the helix axis, the transmitting antenna may be fixed and the helix and ground plane assembly rotated on the helix axis, the axial ratio being the ratio of the maximum to minimum field intensity readings. The latter method was used to obtain the axial ratio *vs.* frequency curve of Fig. 12 measured in the direction of the axis of the helix of Fig. 5. Above about 300 Mc the axial ratio is nearly constant at about 1.2 while below 300 Mc it rises sharply.

The measured current distribution along the helix of Fig. 5 is presented in Fig. 13 for 250 and 450 Mc. It is to be noted that the two distributions are of distinctly different form. The distribution at 250 Mc is characteristic of those measured below 300 Mc while that at 450 Mc is typical of those measured above 300 Mc where the helix is radiating in the axial mode. The 250

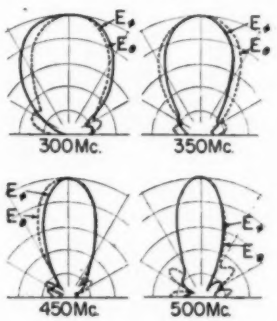


FIG. 7. Measured E_ϕ and E_θ patterns (relative field intensity, as function of ϕ, in x–z plane) for helical antenna of Fig. 5 radiating in axial mode.

Mc distribution is nearly a pure standing wave which can be interpreted as the resultant of two waves of equal amplitude traveling in opposite directions. The distribution at 450 Mc, while showing fluctuations, is much more uniform. A study of this distribution suggests that it is the resultant of four waves having the current distributions illustrated in Fig. 14. There are 2 waves (1 and 2) traveling toward the open end. Wave 1 is exponentially attenuated while wave 2 is propagated with little or no attenuation. At the open end, wave 2 is reflected giving rise to two waves (3 and 4) in the reverse direction. Wave 3 is exponentially attenuated while 4 is propagated. Since wave 4 is smaller than 2, the current minima are not zero in the central helix region. The observed standing wave ratio (*SWR*) in this region of the helix may be taken as

$$SWR=\frac{I_2+I_4}{I_2-I_4}, \tag{3}$$

where I_2=current magnitude of wave 2, and I_4=same for wave 4. The *SWR* in Fig. 13 (lower) is about 1.5 so that $I_4=0.2\ I_2$. Since wave 4 is small compared to wave 2 and since waves 1 and 3 are confined to the ends, it is possible, as shown later, to calculate the approximate radiation pattern of a long helix on the assumption that only a single traveling wave (2) is present.

From symmetry it appears that when one traveling wave is present on a helix of a large number of turns, the radiation in the direction of the axis (if not zero) is circularly polarized regarless of the helix dimensions. This was first suggested to the authors by Dr. George Sinclair. If the axial ratio in the direction of the axis is found to exceed unity, it indicates the presence of other waves on the helix which produce a

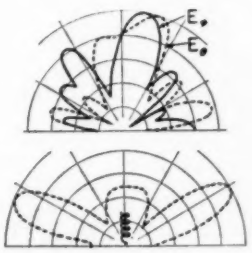


FIG. 8. Measured patterns of helices radiating in multi-lobed and conical modes.

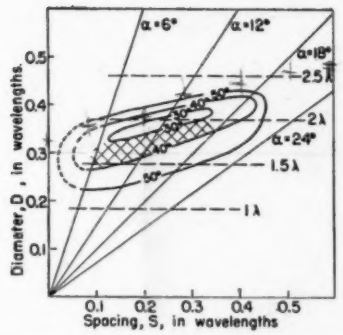


FIG. 9. Contours of measured half-power beam-width of E_ϕ patterns for helices of constant physical length as a function of D_λ and S_λ. Cross-hatched area is region in which side-lobe level is a minimum.

circularly polarized component with an *E*-vector rotation opposite to that of the principal circularly polarized component. The observed axial ratio on the axis is usually slightly greater than unity. Confining our attention to waves 2 and 4, this may be explained by assuming that wave 2 radiates, for example, a right-circularly polarized field while wave 4 radiates (via a reflection from the ground plane) a left-circular field of smaller magnitude.† In general, the axial ratio of an elliptically polarized field having right- and left-circularly polarized components is

$$A.R.=\frac{E_R+E_L}{E_R-E_L}, \tag{4}$$

where E_R=magnitude of right-circular component (assumed larger) and E_L=magnitude of left-circular component (assumed smaller). In the case under consideration, $E_R=kI_2$ and $E_L=kk'I_4$; where I_2=current magnitude of wave 2, I_4=same for wave 4, k=constant of proportionality, and k'=reflection coefficient at ground plane ($0\leqslant k'\leqslant 1$). Thus (4) becomes

$$A.R.=\frac{I_2+k'I_4}{I_2-k'I_4}. \tag{5}$$

† When a left-handed helix is observed from a point on the axis, the *E*-vector of the radiated wave (approaching observer) rotates clockwise as a function of time. This is usually referred to as right-circular polarization as in Max Born, *Optik*, p. 24 (Julius Springer Verlag, Berlin, 1933), and in J. A. Stratton, *Electromagnetic Theory*, p. 280 (McGraw-Hill Book Company, Inc., New York, 1941). Two recent references on properties and applications of circularly polarized radio waves are: W. Sichak and S. Milazzo, "Antennas for circular polarization," paper presented at I.R.E. Convention, Mar. 1947, N. Y.; and T. B. Friedman, "Circularly polarized waves give better f-m service area coverage," Tele-Tech. 6, no. 8, 26 (1947).

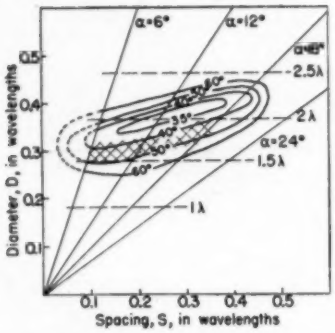


FIG. 10. Contours of measured half-power beam-width of E_θ patterns.

PATTERN CALCULATIONS

A circular helix may be treated in an approximate way by assuming it is of square cross section as in Fig. 15A and carries a uniform traveling wave. The total field from a single turn is then the resultant of the fields of four linear elements. Radiation patterns of helices of circular and square cross section with equal perimeters have been measured and found to be nearly the same. It is believed that the approximation is slightly improved by taking the area of the square helix equal to that of the circular so that $g=(\pi)^{\frac{1}{2}}D/2$ where g is the length of one side of the square cross section. The square helix analysis is, of course, a correct representation for a helix of square cross section. This analysis may be easily extended to the case of a helix of rectangular cross section and this, in turn, approximately to a helix of elliptical cross section by taking the area of the rectangle equal to the area of the ellipse.

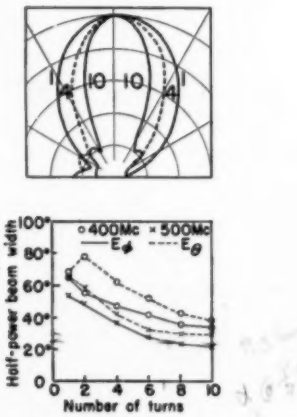


FIG. 11. Measured E_ϕ patterns of 12° helix of 10, 4, and 1 turns (upper) and half-power beam-width variation of E_ϕ and E_θ patterns as a function of the number of turns for two frequencies (lower).

Short Linear Conductors††

Returning to the square helix as used to analyze one of circular cross section, a first step is to obtain an expression for the radiation pattern of a uniform traveling wave on a short linear conductor, such as element 1 in Fig. 15A. Alford[8] obtained this expression for the special case where the phase velocity on the conductor equals the velocity of light. Since, in general, the phase velocity on the helix is different from the velocity of light, a more general solution is required. Referring to Fig. 16, a linear conductor of length b is placed coincident with the z-axis of a cylindrical coordinate system with coordinates ρ, ξ, and z. A point on the conductor is designated z_1. The H_ξ field††† component at the y–z plane is[9,10]

$$H_\xi = j\omega\epsilon \operatorname{curl} \Pi_z = -j\omega\epsilon(\partial \Pi_z/\partial\rho), \qquad (6)$$

where Π_z is the z-component of the Hertz vector

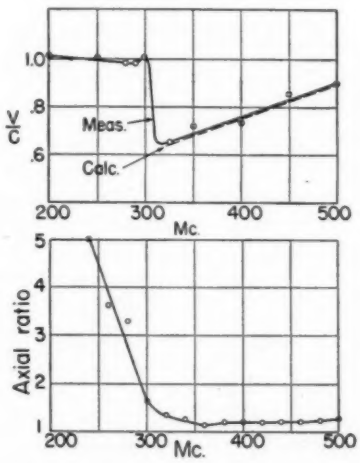


FIG. 12. Measured and calculated phase velocity ratio and measured axial ratio for helix of Fig. 5 as functions of frequency.

†† Although the following discussion is primarily concerned with conductors less than one wave-length long, the expressions may be applied to linear conductors of any finite length with a uniform traveling wave.

[8] Andrew Alford, "A discussion of methods employed in calculations of electromagnetic fields of radiating conductors," Elec. Comm. **15**, 70(1936).

††† Rationalized MKS units are used.

[9] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 456.

[10] J. Aharoni, *Antennae* (Oxford University Press, London, 1946), p. 24.

given by

$$\Pi_z=\frac{1}{4\pi j\omega\epsilon}\int_0^b\frac{[i]}{r}dz_1, \tag{7}$$

and where $[i]$ is the retarded value of the current on the conductor given by

$$[i]=i_0\sin\omega\left(t-\frac{r}{c}-\frac{z_1}{v}\right), \tag{8}$$

where $v=pc$. Let $u=t-r/c-z_1/pc$. Then,

$$\Pi_z=\frac{i_0c}{4\pi j\omega\epsilon}\int_{u_1}^{u_2}\frac{\sin\omega u}{z-z_1-r/p}du, \tag{9}$$

and at a large distance, r, compared to b,

$$H_\xi=\frac{i_0p\rho}{4\pi r}\frac{\sin\omega u_2-\sin\omega u_1}{z-z_1-r/p}. \tag{10}$$

By trigonometric rearrangement, and introducing also $E_\beta=(\mu/\epsilon)^{\frac{1}{2}}H_\xi$, (10) becomes

$$E_\beta=\frac{i_0p\mu^{\frac{1}{2}}}{2\pi r\epsilon^{\frac{1}{2}}}\left\{\frac{\sin\beta}{1-p\cos\beta}\sin\left[\frac{\omega b}{2pc}(1-p\cos\beta)\right]\right\}$$
$$\angle\left[\omega\left(t-\frac{r_1}{c}\right)-\frac{\omega b}{2pc}(1-p\cos\beta)\right]. \tag{11}$$

When $p=1$, (11) is equivalent to the expression developed by Alford. The factor in braces { } gives the amplitude variation of E_β as a function of β and b. Phase, as given by the factor $\angle[\ ]$, is referred to the origin of the coordinates in Fig. 16. A plot of E_β *vs.* β is the radiation pattern of E_β for a conductor of length b.

Patterns calculated from (11) for linear conductors 0.1, 0.25, 0.5, and 0.75 wave-lengths long are presented in Fig. 17 for $p=1$. All are adjusted to the same maximum amplitude. The lower half of the patterns is not shown. As the length, b, increases, the beam-width of the pattern decreases, and also the direction of the maximum tilts in the direction of the traveling wave on the conductor. Referring to Fig. 15C, the tilt-angle between the normal to the conductor and the direction of the maximum is τ. In Fig. 18 the half-power beam-width and tilt-angle, τ, for linear conductors of lengths up to 0.8 wave-lengths are shown for $p=1$ and $p=0.8$. It appears that a reduction in phase velocity as given by

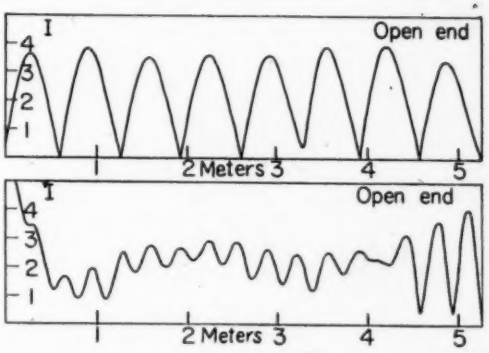


FIG. 13. Measured current distributions on helix of Fig. 5 at 250 (upper) and 450 Mc (lower).

$p=0.8$ is similar in effect to a small increase in the length of the conductor. When $\alpha=\tau$ the maximum of the radiation pattern coincides with the direction of the helix axis. Curves for $\alpha=\tau$ and $p=1$ and $p=0.8$ are presented on a D-S graph in Fig. 19.

Fields from parallel segments of the helix arrive in-phase at a distant point on the axis when $L=p(S+\lambda)$. This condition is not the one giving the sharpest pattern but approximates it. For a helix $L=[(\pi D)^2+S^2]^{\frac{1}{2}}$ and

$$D_\lambda=[p^2(S_\lambda+1)^2-S_\lambda^2]^{\frac{1}{2}}/\pi. \tag{12}$$

When $p=1$, (12) reduces to (1). Three curves (A) for the in-phase condition given by (12) are presented in Fig. 19 for $p=0.8$, 0.9, and 1.0. It is significant that the region of the D-S graph where these curves and the two $\alpha=\tau$ curves are in close proximity corresponds to the observed axial mode region as indicated by the contours of Fig. 9.

Helix of One Turn

Referring to Fig. 15, the equation for the ϕ component ($E_{\phi 1}$) of the E-vector at a large distance in the x–z plane due to element 1 of the square helix is obtained by putting $\beta=3\pi/2$

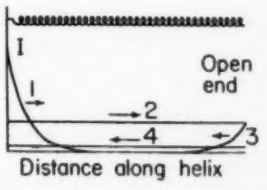


FIG. 14. Resolution of total current distribution into current distributions of four component waves.

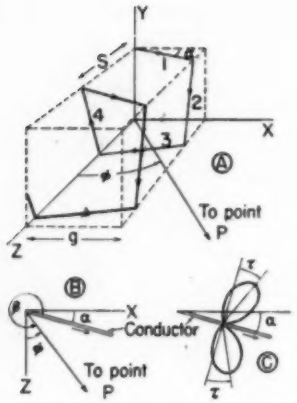


FIG. 15. Helix of square cross section used in analysis.

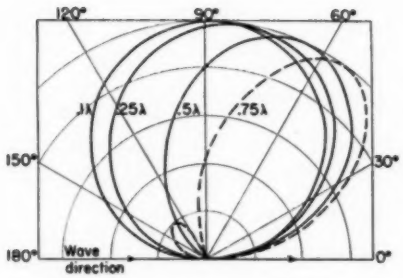


FIG. 17. Calculated relative field intensity patterns produced by a wave traveling with velocity of light on linear conductors 0.1, 0.25, 0.5, and 0.75 wave-lengths long.

$+\alpha+\phi$ and $b=g/\cos\alpha$ in (11). When $t=0$ we have

$$E_{\phi 1}=k\frac{\sin\beta}{A}\sin BA\angle\left(-\frac{\omega r_1}{c}-BA\right), \quad (13)$$

where

$$k=\frac{i_0 p\mu^{\frac{1}{2}}}{2\pi r\epsilon^{\frac{1}{2}}}, \quad A=1-p\cos\beta, \quad \text{and} \quad B=\frac{\omega g}{2pc\cos\alpha}.$$

The expressions for E_ϕ due to elements 2, 3, and 4 (Fig. 15A) are obtained in a similar way. The resultant ϕ-component ($E_{\phi T}$) resulting from a single turn square helix is then the sum of the fields of the four elements, or

$$E_{\phi T}=k\frac{\sin\beta}{A}\sin BA\angle\left(-BA-\frac{\omega r_1}{c}\right)$$
$$+k\frac{\sin\beta''\sin BA''\sin\alpha\sin\phi}{A''(1-\sin^2\alpha\cos^2\phi)^{\frac{1}{2}}}\angle\left[-BA''-\frac{L\omega}{4pc}+\frac{\omega}{c}\left(\frac{S\cos\phi}{4}+g\sin\phi-r_1\right)\right]$$
$$+k\frac{\sin\beta'}{A'}\sin BA'\angle\left[-BA'-\frac{L\omega}{2pc}+\frac{\omega}{c}\left(\frac{S\cos\phi}{2}+g\sin\phi-r_1\right)\right]$$
$$+k\frac{\sin\beta''\sin BA''\sin\alpha\sin\phi}{A''(1-\sin^2\alpha\cos^2\phi)^{\frac{1}{2}}}\angle\left[-BA''-\frac{3L\omega}{4pc}+\frac{\omega}{c}\left(\frac{3S\cos\phi}{4}-r_1\right)\right], \quad (14)$$

where

$$\beta=3\pi/2+\alpha+\phi, \qquad A=1-p\cos\beta,$$
$$\beta'=\pi/2-\alpha+\phi, \qquad A'=1-p\cos\beta',$$
$$\beta''=+\arccos(\sin\alpha\cos\phi), \qquad A''=1-p\cos\beta'',$$

In (14) L is the length of a single turn of the helix whose pattern is being calculated. For a circular helix $L=\pi D/\cos\alpha$. If the contributions of elements 2 and 4 are neglected, which is a good approximation when both α and ϕ are small, the pattern calculation is considerably simplified. Making this approximation, putting $k=1$, $r_1=$constant, and introducing $g=\pi^{\frac{1}{2}}D/2$ for a circular helix,

$$E_{\phi T}=\frac{\sin\beta}{A}\sin BA\angle(-BA)+\frac{\sin\beta'}{A'}\sin BA'$$
$$\angle[-BA'-2\pi^{\frac{1}{2}}B$$
$$+\pi(S_\lambda\cos\phi+\pi^{\frac{1}{2}}D_\lambda\sin\phi)] \quad (15)$$

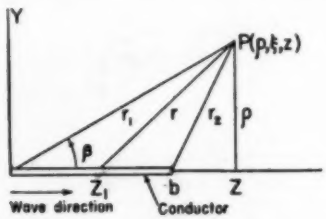


FIG. 16. Cylindrical coordinates for short linear conductor calculations.

where

$$B=\omega g/2pc\cos\alpha=D_\lambda\pi^{\frac{1}{2}}/2p\cos\alpha.$$

Equation (15) is the approximate expression for the E_ϕ pattern at a large distance in the x–z plane caused by a traveling wave on a circular helix of one turn and applies particularly when either α, ϕ, or both are small.

In the case of the E_θ component, only elements 2 and 4 contribute so that the field of a single turn is

$$E_{\theta T}=E_{\theta 2}+E_{\theta 4}$$

and

$$E_{\theta T}=k\frac{\sin\beta''\sin BA''\cos\alpha}{A''(1-\sin^2\alpha\cos^2\phi)^{\frac{1}{2}}}\angle\left[-BA''-\frac{\omega r_1}{c}\right]$$
$$-k\frac{\sin\beta''\sin BA''\cos\alpha}{A''(1-\sin^2\alpha\cos^2\phi)^{\frac{1}{2}}}\angle\left[-BA''-\frac{L\omega}{2pc}\right.$$
$$\left.+\frac{\omega}{c}\left(\frac{S\cos\phi}{2}-g\sin\phi-r_1\right)\right] \quad (16)$$

where $\beta''=\arccos(\sin\alpha\cos\phi)$ and $A''=1-p\cos\beta''$. Putting $k=1$ and introducing $g=\pi^{\frac{1}{2}}D/2$ for a circular helix, the approximate θ pattern of a single turn is

$$|E_{\theta T}|=2\frac{\sin\beta''\sin BA''\cos\alpha}{A''(1-\sin^2\alpha\cos^2\phi)^{\frac{1}{2}}}$$
$$\times\sin\tfrac{1}{2}[\pi(S_\lambda\cos\phi-\pi^{\frac{1}{2}}D_\lambda\sin\phi)-2\pi^{\frac{1}{2}}B]. \quad (17)$$

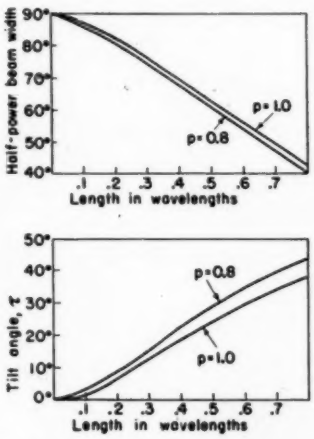


FIG. 18. Half-power beam-width and tilt-angle, τ, of radiation patterns produced by a traveling wave on linear conductors of lengths up to 0.8 wave-lengths for two values of phase velocity factor p.

Array Factor

Equations (15) and (17) give the approximate patterns of E_ϕ and E_θ for a single turn circular helix. The field for a helix of n turns can be calculated as an array of such turns. This calculation is facilitated by first evaluating an array amplitude factor Y for any array of n identical non-directional point sources of equal amplitude spaced a distance, S, apart. The angle, ϕ, is measured from the axis or line of sources. When the number of sources is even (n_e) we have,

$$Y_{n_e}=\frac{2}{n_e}\left|\sum_{n=2}^{n_e}\cos\left(\frac{n-1}{2}\psi\right)\right|$$

where $n=2, 4, 6$, etc., (18)

and for an odd number of sources (n_o),

$$Y_{n_o}=\left|\frac{1}{n_o}+\frac{2}{n_o}\sum_{n=3}^{n_o}\cos\left(\frac{n-1}{2}\psi\right)\right|$$

where $n=3, 5, 7$, etc. (19)

In general, the auxiliary function $\psi=2\pi S_\lambda\cos\phi+\gamma$, where γ is the phase difference between sources. In the case of a helix, assuming a uniform traveling wave along the helical conductor, $\gamma=-\omega L/pc$, and

$$\psi=2\pi(S_\lambda\cos\phi-L_\lambda/p), \quad (20)$$

where $L_\lambda=\pi D_\lambda/\cos\alpha$ for a circular helix. The array amplitude factor Y for 1 to 11 sources *vs.* the auxiliary function, ψ, is shown by the family of curves in Fig. 20. The curves are evaluated from (18) and (19). As applied to a helix, these curves correspond to 1 to 11 turns. If ψ is expressed as a function of ϕ by (20), the array amplitude factor as a function of ϕ for a helix of n turns may be obtained directly from Fig. 20. This factor multiplied by the amplitude factor

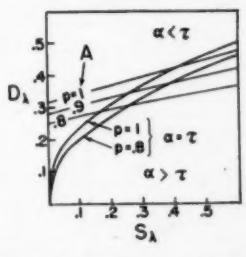


FIG. 19. Curves for $\alpha=\tau$ and for in-phase fields on axis (A) compared on a $D-S$ chart for helices.

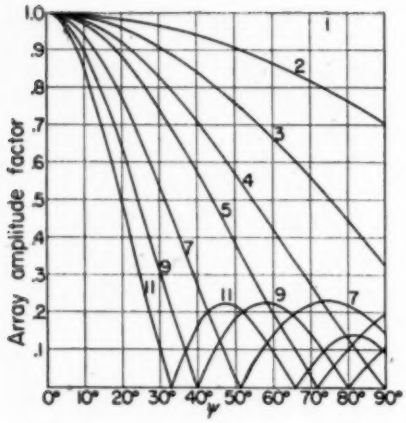


FIG. 20. Array amplitude factor *vs.* auxiliary function ψ for arrays of 1 to 11 turns.

of a single turn then yields the pattern of the entire helix or

$$|E_\phi(\phi)| = |E_{\phi T}(\phi)|\, Y_n(\phi), \tag{21}$$

and

$$|E_\theta(\phi)| = |E_{\theta T}(\phi)|\, Y_n(\phi), \tag{22}$$

where

$|E_\phi(\phi)|$ = amplitude of E_ϕ as a function of ϕ,
$|E_{\phi T}(\phi)|$ = amplitude of E_ϕ for one turn as a function of ϕ,
$Y_n(\phi)$ = array amplitude factor for n turns as function of ϕ, etc.

The effect of the ground plane on the patterns is small and is neglected.

Comparison of Measured and Calculated Patterns

Following the above procedure the approximate patterns of a typical helix have been calculated and found to be in good agreement with measured patterns. For example, the measured E_ϕ and E_θ patterns are compared in Fig. 21 for the helix of Fig. 5 with the calculated patterns

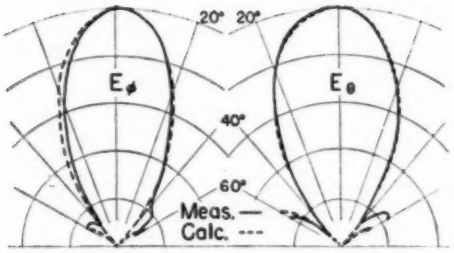


FIG. 21. Calculated and measured E_ϕ and E_θ patterns (as function of ϕ, in x–z plane) of 12°, 7-turn helix shown in Fig. 5 at 450 Mc.

using (15) and (17). In this example, the amplitude of $E_{\theta T}(\phi)$ is more constant (less directional) than $E_{\phi T}(\phi)$. Since the array factor is the same for both E_θ and E_ϕ, the resultant E_θ pattern of the helix is not as sharp as the E_ϕ pattern.

The quantities which must be known to calculate a helix pattern by the above method are D_λ, S_λ or α, p, and n. All but p are known from the physical dimensions of the helix and the frequency. Pending a method for calculating p from the helix geometry, it must be measured either directly, as for Fig. 12, or indirectly by observing the value of ϕ for the first null of the measured radiation pattern. By the indirect method ψ at this value of ϕ can be taken from Fig. 20 and both ψ and ϕ substituted into (20), giving the value of p. From the null angles of the measured patterns of the helix of Fig. 5 at various frequencies, p was found in this manner in the frequency range of the axial mode. The variation is approximately a straight-line relation and is presented as the calculated curve in Fig. 12.

ACKNOWLEDGMENT

The authors wish to thank Dr. George Sinclair of the Antenna Laboratory, Ohio State University, for his interest and very helpful discussions.

On the Structure of Cellulose Acetate Molecules in Acetone Solution[1]

SEYMOUR J. SINGER[2] AND HERMAN MARK
Institute of Polymer Research, Polytechnic Institute of Brooklyn, Brooklyn, New York

(Received July 30, 1947)

Data obtained from sedimentation velocity and viscosity measurements on a series of cellulose acetate fractions in acetone are interpreted in terms of two models, the impenetrable ellipsoid of rotation and the uniform density random coil. The lack of agreement using the former model is shown to be a real effect, rather than one determined by experimental inaccuracy. As other investigations have revealed, the uniform density random coil is not a satisfactory model for these molecules either. Probably the best hydrodynamical model for these and other extended but not perfectly rigid molecules in solution is a penetrable ellipsoid of rotation.

A. INTRODUCTION

IN recent years, combination of results from many physico-chemical measurements has yielded much valuable data on the size of large molecules in solution. Accurate molecular weight determinations are readily available. Concerning the macrostructures, or "shapes," of these molecules, however, there is considerably less certainty in interpreting the data. As yet, only in those cases where the shapes of the molecules are well defined, i.e., when the molecules are non-solvated rigid spheres or rods, can indirect methods like sedimentation and viscosity be used unequivocally. By utilizing the model of a rigid ellipsoid of rotation for the rod-like molecules, viscosity data may be interpreted by the equation of Simha,[3] and sedimentation and diffusion data by that of Herzog, Illig, and Kudar,[4] or equivalently of Perrin,[5] and axis ratios of these ellipsoids calculated.

For many protein solutions, such a model is probably satisfactory in order to describe their hydrodynamical behavior. In the case of dilute, buffered tobacco mosaic virus solutions, the axis ratios obtained for these rod-like particles by these indirect measurements agree with the value obtained by the electron microscope,[6] but the degree of solvation of these particles in solution makes the results somewhat ambiguous.

Difficulties arise, however, when the molecules are not rigid and compact, but assume a variety of energetically equivalent configurations under the influence of thermal agitation, as in the case of many polymer molecules in solution. Here we consider a time-average "shape," and the problems of the degree of swelling and the permeability of the swollen solute mesh to solvent become significant.

First, however, we may divide polymer solutions into two somewhat overlapping categories for our purposes, those in which the solute molecules are relatively rigid and elongated, such as solutions of cellulose, cellulose derivatives, and related compounds, and those in which the solute molecules are flexible and highly coiled, like polyisobutylene. In solutions of the former type, the observed large viscosities are due principally to the considerable asymmetry of the solute molecules, while in the solutions of the latter type, the large viscosities may be attributed to two factors: (1) the large volume fraction occupied by the swollen spheres of solute, and (2) the frictional resistance offered to the motion of the solvent through that part of the solute mesh which is penetrable. Many investigators have developed relations which ascribe the viscosity of solutions of coiled polymers to the swelling of the solute molecules,[7] while theories which emphasize the frictional resistance met by

[1] Presented before a meeting of the American Physical Society, Columbia University, New York, January 1947.
[2] Du Pont Research Fellow, 1946–1947. This paper is from the thesis submitted by S. J. Singer in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Polytechnic Institute of Brooklyn, June 1947.
[3] R. Simha, J. Phys. Chem. **44**, 25 (1940). This equation has come to be regarded as the most useful of those so far put forward.
[4] R. O. Herzog, R. Illig, and H. Kudar, Zeits. f. physik. Chemie **A167**, 329 (1933).
[5] F. Perrin, J. de phys. et rad. **7**, 1 (1936).
[6] M. Lauffer, J. Am. Chem. Soc. **66**, 1188 (1944).
[7] See, for example, the review article by R. H. Ewart, in *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. II, p. 197.

solvent streaming through the coiled molecule are to be found in the papers of Kuhn and Kuhn,[8] Huggins,[9] Hermans,[10] Kramers,[11] and Debye.[12] More recently, however, Debye[13] has developed a theory for the viscosity of solutions of coiled molecules which essentially takes into account both of these factors, and determines to what extent each is operative.

In the following sections, we shall discuss viscosity and sedimentation data obtained with fractions of cellulose acetate in acetone. From information previously derived by a variety of methods, it is known that these solutions fit more readily into the first category mentioned above, that is, the solute molecules are relatively elongated in solution. In interpreting our data, therefore, we shall attempt to determine how well the models of the rigid ellipsoid of rotation and of the penetrable random coil fit the experimental results, and we shall not be concerned with the model of a coil highly swollen with, and practically impermeable to, solvent.

B. THE ELLIPSOID MODEL

Simha[14] has considered the applicability of the model of an impenetrable ellipsoid of rotation to polymer molecules in solution, as a rough approximation to their actual character. Being relatively rigid extended molecules, cellulose acetate particles in a fairly good solvent like acetone might be expected to conform to this model more closely than most other polymers. Because of the polydispersity of some of the cellulose derivatives previously examined by sedimentation and viscosity measurements,[15–17] and the chemical inhomogeneity of others,[18] it hasn't been possible clearly to determine how well the ellipsoid model fits the data.

One of us has performed sedimentation and diffusion experiments with four cellulose acetate fractions in acetone.[19] Three of these fractions were products of the fractionation carried out by Sookne, Rutherford, Mark, and Harris[20] and analyzed for the same acetyl content, 38.6 percent. They were investigated by Sookne and Harris[21] with viscosity and osmotic pressure measurements. The fourth fraction was prepared by W. Badgley[22] using a similar polymer and fractionation technique, but was not analyzed. The molecular weights obtained by sedimentation and diffusion are close to those obtained from osmotic pressure measurements, indicating a high degree of homogeneity of molecular weights in the fractions.

By means of the Perrin and Simha equations, which have been shown by Lauffer to agree in tobacco mosaic virus experiments for an axis ratio of about 20:1, we have evaluated our sedimentation and the Sookne and Harris viscosity data. The Perrin equation for a prolate* ellipsoid is:[5]

$$\frac{f}{f_0}=\frac{(1-\rho^2)^{\frac{1}{3}}}{\rho^{\frac{1}{3}}\log\dfrac{1+(1-\rho^2)^{\frac{1}{2}}}{\rho}} \qquad \rho<1 \qquad (1)$$

in which f is the experimental frictional coefficient per mole of solute, the proportionality constant between the force of resistance per mole, and the velocity of sedimentation of the particles; f_0 the hypothetical frictional coefficient of the spherical particles equivalent to the ellipsoidal ones under consideration; and ρ the ratio of the short and long axes of the ellipsoid of rotation.

The experimentally observed frictional coefficient is evaluated from the sedimentation con-

[8] W. Kuhn and H. Kuhn, Helv. Chim. Acta **26**, 1394 (1943); **28**, 1533 (1945); **29**, 71 (1946).
[9] M. L. Huggins, J. Phys. Chem. **42**, 911 (1938); **43**, 439 (1939).
[10] J. J. Hermans, Physica **10**, 777 (1943).
[11] H. A. Kramers, J. Chem. Phys. **14**, 415 (1946).
[12] P. Debye, J. Chem. Phys. **14**, 636 (1946).
[13] P. Debye, in a paper presented before the meeting of the American Physical Society at Columbia University, New York, January 1947.
[14] R. Simha, J. Chem. Phys. **13**, 188 (1945).
[15] N. Gralén, Dissertation, Upsala, 1944.
[16] H. Mosimann, Helv. Chim. Acta **26**, 61 (1943).
[17] I. Jullander, Arkiv. f. Kemi. Min. Geol. **A21**, 1 (1945).
[18] R. Signer and P. v. Tavel, Helv. Chim. Acta **21**, 535 (1938).
[19] S. J. Singer, J. Chem. Phys. **15**, 341 (1947).
[20] A. Sookne, H. Rutherford, H. Mark, and M. Harris, J. Research Nat. Bur. Stand. **29**, 123 (1942).
[21] A. Sookne and M. Harris, Ind. Eng. Chem. **37**, 475 (1945).
[22] W. Badgley, Ph.D. thesis, Polytechnic Institute of Brooklyn, 1945.

* The model of an impenetrable *oblate* ellipsoid of rotation leads to particle dimensions which are impossible. That is, for a given axis ratio prolate ellipsoids create a considerably greater frictional ratio (and viscosity increment) than oblate ones, and in order to account for the large ratios obtained with cellulose derivatives in solution, one would require oblate ellipsoids with such a large axis ratio that one of the particle dimensions would become impossibly small.

TABLE I. Sedimentation and viscosity data on cellulose acetates.

Fraction	Molecular weight Osm.	Molecular weight Ultracent.	f/f_0	$[\eta]$	$[\eta]_v$	Axis ratios Visc.	Axis ratios Ultracent.
15	11,000	10,350	2.22	0.23	34	18	26
9	53,000	51,000	3.50	1.37	201	54	73
2	130,000	143,000	4.78	2.75	404	80	140
2c	—	194,000	5.10	3.05	449	86	159

TABLE II. Dimensions of cellulose acetate molecules.*

Fraction	Extended length	R (calculated)	Ellipsoid dimensions Viscosity l**	d	Ultracent. l	d	Random coil dimensions Viscosity $2R_0$	L	Ultracent. $2R_0$	L
15	190	50	190	10.7	250	9.5	125	19	90	10
9	930	110	680	12.7	840	11.4	390	58	240	28
2	2600	180	1250	15.8	1820	13.0	690	105	450	54
2c	3500	210	1460	17.0	2200	13.8	790	120	540	64

* All dimensions in angstrom.
** l is the major, d the minor, axis of the ellipsoid.

stant extrapolated to infinite dilution,[19] s_∞, by the equation:

$$f = M(1-\bar{V}d)/s_\infty. \tag{2}$$

M is the molecular weight of the particle of partial specific volume $\bar{V}$ sedimenting in a medium of density d. The frictional coefficient of the equivalent sphere is obtained from Stokes' Law as follows:

$$f_0 = 6\pi\eta N_A\left(\frac{3M\bar{V}}{4\pi N_A}\right)^{\frac{1}{3}}, \tag{3}$$

where η is the viscosity of the solvent and N_A is Avogadro's number.

The equation of Simha[3] for the viscosity increment resulting from prolate ellipsoids of rotation in solution is:

$$[\eta]_v = \frac{(1/\rho)^2}{15\log(2/\rho)-45/2} + \frac{(1/\rho)^2}{5\log(2/\rho)-5/2} + \frac{14}{15} \quad \rho<1, \tag{4}$$

in which $[\eta]_v$ is the limiting value, as the concentration of solute approaches zero, of the specific viscosity, η_{sp}, divided by the concentration c_v (expressed as the volume fraction of solute). This limiting value is obtained in the usual manner[23] by plotting η_{sp}/c against c and extrapolating to $c=0$. Generally the experimental concentration unit is expressed in grams of solute per 100 cc of solution, so that the value of $[\eta]$ obtained experimentally must be multiplied by $100/\bar{V}$ in order to apply Eq. (4). This value is referred to as $[\eta]_v$.

The sedimentation and diffusion data are tabulated and discussed in a previous communication.[12] The value of M used in Eq. (2) does not create any significant problem in this investigation, and the molecular weights derived from the Svedberg velocity ultracentrifuge equation were used consistently. For these acetone soluble cellulose acetates, Kraemer's value of $\bar{V}=0.68$ was employed.[24] The values of the axis ratios calculated from Eqs. (1) and (4) are given in Table I.

The discrepancies between the axis ratios obtained from viscosity and ultracentrifuge measurements on the higher molecular weight cellulose acetates are outside the range of experimental error. Two other obvious possible causes for the differences, namely, the low values of M used in Eq. (2), and the orientation of the solute during transport, may also be eliminated. For if the molecular weights substituted in the equation are too low, increasing them only augments the discrepancies, since the axis ratios calculated from the sedimentation data are increased. Concerning the second point, experiments by Badgley and Mark[25] on a high molecular weight fraction of cellulose acetate in acetone demonstrated that increasing the pressure driving the solutions through the viscosimeter capillary did not alter $[\eta]$, and orientation during sedimentation is almost always negligible[26] in polymer solutions.

Another possible source of difficulty arises from the use of $\bar{V}$ in Eq. (2), and to convert $[\eta]$ to $[\eta]_v$. By thus employing $\bar{V}$ to represent the reciprocal of the particle density, one ignores the effect of the solvent which may move with the macromolecule as a kinetic unit. However, while this effect is a serious one for flexible polymer molecules in solution, it does not represent a severe assumption for these elongated macromolecules.

[23] See, for example, M. L. Huggins, in *High Polymers* (Interscience Publishers, Inc., New York, 1943), Vol. V, p. 951.

[24] E. O. Kraemer and W. D. Lansing, J. Phys. Chem. **39**, 153 (1935).

[25] W. Badgley and H. Mark, J. Phys. Coll. Chem. **51**, 58 (1947).

[26] S. J. Singer, J. Polymer Sci., in press.

Moreover, since the Perrin and Simha equations have apparently been shown to agree at a relatively small axis ratio, it seems plausible that with more elongated ellipsoids, approximations in both equations would become better, and the agreement should continue.

The conclusion appears to be, then, that the impenetrable prolate ellipsoid model is only qualitatively satisfactory for cellulose acetate in acetone in the molecular weight range 10,000–200,000 which we have studied. It is of interest that similar results have been obtained by Signer and Tavel[18] with methyl celluloses in water, and by Dombrow and Beckmann[27] with amylose acetates in methyl acetate. In these investigations, also with extended molecules in solution, the values of the axis ratios obtained from Perrin's equation are all higher than those obtained from Simha's equation, and the discrepancies increase with increasing elongation of the molecules, in a fashion exactly analogous to the results recorded in Table I. At the present time, it is evident that quantitative applications of axis ratios obtained by these methods are not justified, unless the values derived independently are in agreement.

C. THE RANDOM COIL MODEL

There is, of course, a strong similarity between the processes of sedimentation (or diffusion) and viscosity. Corresponding to parameters obtained from one measurement there will generally be found similar though independent quantities from the other. For example, as was evident in the last section, f and $[\eta]$ are closely connected, and the Perrin and Simha equations demonstrate the relation.

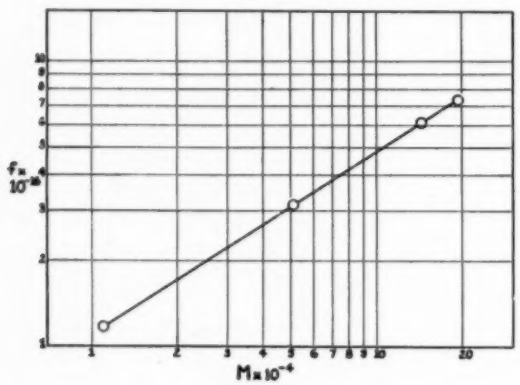


FIG. 1.

In another direction, this connection between f and $[\eta]$ may be utilized. The relation between $[\eta]$ and molecular weight is generally given by the following empirical equation:[28]

$$[\eta]=KM^a, \tag{5}$$

in which K and a are constants. Over a considerable range of molecular weights, however, this equation does not satisfy the data.[25] Recently, Debye[13] has shown that, utilizing the random coil as a model for the polymer molecule in solution, Eq. (5) may be interpreted to give information concerning the macrostructure of the molecule. The constant a in Eq. (5) is not truly a constant, but for any homologous series of random coils it will vary from the value 1 at very low molecular weights (Staudinger's rule) to the value $\frac{1}{2}$ in the limit of high molecular weights; for any particular intermediate range of particle weights Eq. (5) will be satisfied with some value of a between 1 and $\frac{1}{2}$.

In particular, he proposed the replacement of the empirical equation (5) by the theoretical equation

$$[\eta]=[N_A\Omega_s\phi(R_s/L)]/M. \tag{6}$$

Here, R_s is the radius of the spherical volume Ω_s over which the polymer molecule extends with uniform density, L a length which determines the depth to which a disturbance from the outside can enter into the polymer sphere, and the function ϕ, the volume factor, is a known function of the shielding ratio, R_s/L.

In an analogous manner, the empirical equation

$$f=AM^b, \tag{7}$$

in which A and b are constants corresponding but not equal to K and a, respectively, has been interpreted by Debye and Bueche[29] in terms of the random coil model, in which the solute segments uniformly occupy the polymer sphere. Again the value of b varies in any system of randomly coiled molecules from the value 1 for

[27] A. Dombrow and C. O. Beckmann, J. Phys. Coll. Chem. **51**, 107 (1947).

[28] H. Mark, *Der feste Korper* (S. Hirzel, Leipzig, 1938), p. 103.

[29] P. Debye and A. M. Bueche, in a paper given at the meeting of the American Chemical Society, Atlantic City, April 1947.

low molecular weights to $\frac{1}{2}$ in the limit of high molecular weights, and instead of the empirical Eq. (7), the corresponding theoretical one is advanced:

$$f = 6\pi\eta N_A R_s \psi(R_s/L), \qquad (8)$$

in which ψ is a known function of the shielding ratio, and the other symbols have their previous significance.

The system cellulose acetate in acetone is not an especially good one to apply to this theory, since from the extended shape of its solute molecules we can be certain the molecules are not random coils in which each monomer unit can rotate freely about the bond to its neighboring unit. Nevertheless, it is of interest to determine the magnitudes of the parameters involved. In the molecular weight range 10,000–200,000, the sedimentation data yield the value $b = 0.60$, as is shown in the $\log f - \log M$ plot given in Fig. 1. From the investigations of Badgley and Mark[25] over a wide range of molecular weights in a very similar cellulose acetate-acetone system, a was determined as 0.82. This represents probably the best value available at present; other values[14,30] having been derived from measurements over a narrower range of molecular weights. The value of the shielding ratio obtained from the ultracentrifuge is, therefore, 4.2, while from viscosity it is 3.3. In Table II are recorded the dimensions obtained from both sets of data, utilizing both the impenetrable prolate ellipsoid and the uniform density random coil models. The extended length calculated from the degree of polymerization and the x-ray length of the glucose unit (5.1A), and the calculated root mean square separation, R, of the ends of the random coil are also given. The latter is obtained from the formula

$$R^2 = \frac{1-p}{1+p} NA^2, \qquad (9)$$

where p is the cosine of the angle between two consecutive links, N is the degree of polymerization, and A is the length of the link. We have chosen[31] $p = \cos 115° = -0.423$, as a probable value.

D. DISCUSSION

Previous physico-chemical investigations on cellulose acetates in acetone[25,32,33] have indicated that in the low molecular weight range the molecules are extended, while beyond about 80,000, the molecules begin to coil back upon themselves. By comparison of the extended lengths and the values for the major axes of the ellipsoids of rotation given in Table II, this conclusion is again forthcoming.

However, it is interesting to note that the data may be as satisfactorily interpreted by a uniform density random coil as by an impenetrable ellipsoid model, with neither giving the desired agreement between sedimentation and viscosity data. From the large values of the axis ratios calculated from the ellipsoid model, and the relatively low values of the shielding ratio (high penetrability) obtained from the random coil model, it would appear that a model which is a combination of the two, that is, a *penetrable* ellipsoid of rotation, would very likely be the most reasonable and useful one for cellulose acetates in acetone, and for other extended but not perfectly rigid macromolecules in solution.

ACKNOWLEDGMENT

The authors are indebted to Dr. K. G. Stern for his help and advice, and to Drs. A. M. Sookne, Milton Harris, and W. J. Badgley for furnishing the fractions used in this study.

[30] A. Bartovics and H. Mark, J. Am. Chem. Soc. **65**, 1901 (1943).

[31] F. T. Peirce, Trans. Faraday Soc. **42**, 545 (1946).

[32] P. Doty and H. S. Kaufman, J. Phys. Chem. **49**, 583 (1945).

[33] R. S. Stein and P. Doty, J. Am. Chem. Soc. **68**, 159 (1946).

The Effect of Ground Constants on the Characteristic Values of the Normal Modes in Non-Standard Propagation of Microwaves

C. L. PEKERIS*
Columbia University, New York, New York
(Received August 25, 1947)

It is shown that for the case of a strongly absorbing ground ($\beta \gg 1$) the characteristic values D of the normal modes are given by

$$D = D_a - (\dot{U}^2/\beta) + [\dot{U}^3(\partial \dot{U}/\partial D)/\beta^2] + \cdots, \quad \text{(a)}$$

where

$$\beta_v = i(k_0/q)^{\frac{1}{2}}(\epsilon_1 - 1)^{\frac{1}{2}}/\epsilon_1; \quad \beta_h = \beta_v \epsilon_1, \quad \text{(b)}$$

D_a is the characteristic value obtained under the usual assumption that the height-gain function U vanishes at the ground, and $\dot{U}$ denotes dU/dz at $z=0$. In the other limit of a perfectly conducting ground ($\beta \ll 1$) the following expansion holds

$$D = D_c + \beta U^2 + \beta^2 U^3(\partial U/\partial D) + \cdots, \quad \text{(c)}$$

where D_c is the characteristic value obtained under the condition $\dot{U}(0)=0$. Since $\dot{U}$ and U in (a) and (c), respectively, are known from the normalization condition, it is possible to determine the first correction terms to the characteristic values in each case.

INTRODUCTION

IN the normal mode theory[1] of propagation of microwaves under non-standard conditions of atmospheric refraction, the characteristic values are derived under the assumption that the height-gain function U vanishes at the ground. This boundary condition is valid only when a parameter β, depending on the ground constants, is large, the so-called case of strong absorption, where

$$\beta_v = i(k_1^2 - k^2)^{\frac{1}{2}}(qk_0^2)^{-\frac{1}{2}}/\epsilon_1 \simeq i(k_0/q)^{\frac{1}{2}}(\epsilon_1 - 1)^{\frac{1}{2}}/\epsilon_1; \quad \beta_h = \beta_v \epsilon_1, \quad (1)$$

$$\epsilon_1 = \epsilon - i(\sigma/\omega); \quad k_0 = 2\pi/\lambda; \quad q = 2/a_e, \quad (2)$$

the subscripts v and h refer to vertical and horizontal polarization, respectively, and a_e denotes the effective radius of the earth. In Figs. 1 and 2 values of β_v^{-1} and β_h^{-1} are given for sea water and "average soil." It is seen that β_h^{-1} is always small, but that β_v^{-1} becomes large at the longer wave-lengths. The purpose of this investigation is primarily to ascertain the change in the characteristic values of the normal modes resulting from the finiteness of β. Such a correction, although small, is of importance under conditions of strong trapping and vertical polarization, where in the first approximation the theoretical horizontal decrement, which is proportional to the imaginary part of D, is extremely small. It turns out that the first correction term can be determined, without additional computation, from the known properties of the height-gain function.

FORMAL SOLUTION

The formal solution for the Hertzian potential Ψ resulting from a dipole radiating in an atmosphere under non-standard refraction conditions can be obtained along the lines followed[2] in the case of standard atmospheric refraction. The wave equation

$$\nabla^2\Psi + k_0^2\mu^2\Psi, \quad k_0 = 2\pi/\lambda, \quad (3)$$

is synthesized from elementary solutions of the form

$$\psi = e^{i\omega t} J_0(kr) F(h) G(k), \quad (4)$$

where, on replacing μ by the modified index of refraction N, F must satisfy the differential equations

$$d^2F/dh^2 + (k_0^2N^2 - k^2)F = 0, \quad h > 0, \quad (5)$$

$$d^2F/dh^2 + (k_1^2N^2 - k^2)F = 0, \quad h < 0, \quad k_1^2 = k_0^2\epsilon_1, \quad \epsilon_1 = \epsilon - i\sigma/\omega. \quad (6)$$

* This paper is based on work done for the Naval Research Laboratory under contract with the Navy's Office of Research and Inventions. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University

[1] Radiation Laboratory Report 680 (1945) by W. H. Furry; C. L. Pekeris, J. Acous. Soc. Am. **18**, 304 (1946); Phys. Rev. **70**, 518 (1946); J. App. Phys. **17**, 1108 (1946).

[2] C. L. Pekeris, J. App. Phys. **18**, 667 (1947).

Let $M(h)$ and $N(h)$ denote the solutions of (5) which at great heights, h, represent downcoming and outgoing waves, respectively, and let $P(h)$ denote the solution of (6) which does not become infinite at great depth. If h_1 denote the height of the transmitter, then the appropriate solutions are

$$F_1(h)=C(k)N(h), \qquad h>h_1, \qquad (7)$$

$$F_2(h)=A(k)M(h)+B(k)N(h), \quad 0<h<h_1, \qquad (8)$$

$$F_3(h)=E(k)P(h), \qquad h<0. \qquad (9)$$

The functions A, B, C, and E of k are determined from the boundary conditions:

$$F_1=F_2; \quad F_2'-F_1'=2k, \quad \text{at } h=h_1, \qquad (10)$$

$$\gamma F_2=F_3, \quad F_2'=F_3',$$

$$(\gamma_v=1/\epsilon_1, \quad \gamma_h=1), \quad \text{at } h=0, \qquad (11)$$

where the prime denotes differentiation with respect to h. It is found that $P(h)$ enters the expression for the potential in the atmosphere only through the combination $P'(0)/P(0)$, for which one can write with very high approximation[3] $i(k_1^2-k^2)^{\frac{1}{2}}$. One thus obtains

$$\Psi_1=-(2e^{i\omega t}/b)\int_0^\infty J_0(kr)kdkN(h)$$

$$\times[M(h_1)+(B/A)N(h_1)], \quad h>h_1, \qquad (12)$$

$$\Psi_2=-(2e^{i\omega t}/b)\int_0^\infty J_0(kr)kdkN(h_1)$$

$$\times[M(h)+(B/A)N(h)], \quad 0<h<h_1, \qquad (13)$$

$$b\equiv MN'-NM'=\text{constant};$$

$$(B/A)=-[\beta M(0)-HM'(0)]/$$

$$[\beta N(0)-HN'(0)], \qquad (14)$$

where β is defined in (1), and $H(=k_0^{-1}q^{-1})$ is the natural unit of height. In terms of $z=h/H$ we have

$$B/A=-[\beta M(0)-\dot{M}(0)]/Q, \qquad (15)$$

$$Q\equiv[\beta N(0)-\dot{N}(0)]; \quad \dot{M}\equiv dM/dz. \qquad (16)$$

The normal mode solution is obtained by evaluating the integrals (12) and (13) in terms of the residues[1] at the zeros k_m in the k-plane of Q:

$$\Psi=-(i\pi e^{i\omega t}/H)\sum_1^\infty H_0^{(2)}(k_m r)U_m(z_1)U_m(z), \qquad (17)$$

where $U_m(z)[=N_m(z)]$ is normalized by the condition

$$\operatorname*{Lim}_{I(k)\to 0^-} -\int_0^\infty U_m^2(z)dz=1, \qquad (18)$$

and must satisfy the boundary condition at the ground

$$\beta U_m(0)-\dot{U}_m(0)=0. \qquad (19)$$

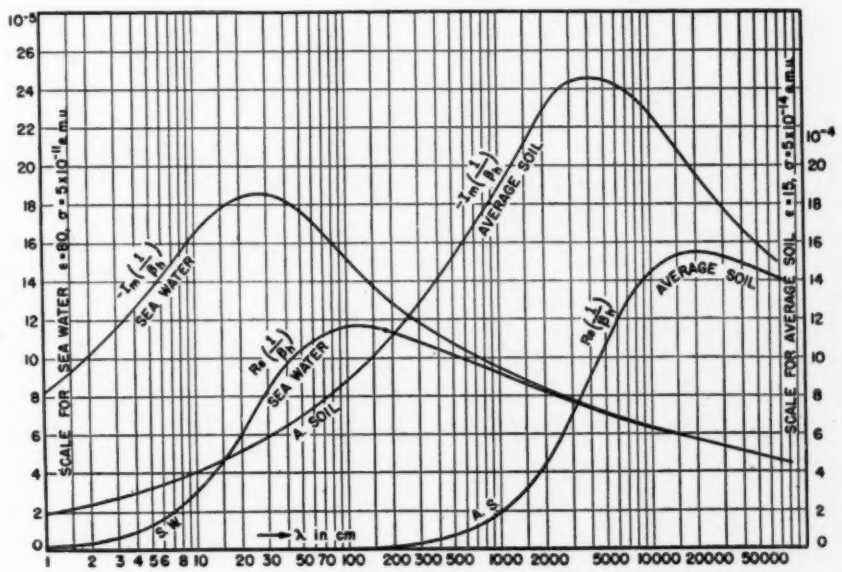


FIG. 1. Values of $(1/\beta_h)$ for "average soil":
$\epsilon=15$, $\sigma=5\times10^{-14}$ e.m.u.
sea water:
$\epsilon=80$, $\sigma=5\times10^{-11}$ e.m.u.

[3] See reference 2, p. 670.

The Effect of Ground Constants on the Characteristic Values of the Normal Modes in Non-Standard Propagation of Microwaves

C. L. PEKERIS*
Columbia University, New York, New York
(Received August 25, 1947)

It is shown that for the case of a strongly absorbing ground ($\beta \gg 1$) the characteristic values D of the normal modes are given by

$$D = D_e - (\dot{U}^2/\beta) + [\dot{U}^2(\partial \dot{U}/\partial D)/\beta^2] + \cdots, \quad \text{(a)}$$

where

$$\beta_v = i(k_0/q)^{\frac{1}{2}}(\epsilon_1 - 1)^{\frac{1}{2}}/\epsilon_1; \quad \beta_h = \beta_v \epsilon_1, \quad \text{(b)}$$

D_e is the characteristic value obtained under the usual assumption that the height-gain function U vanishes at the ground, and $\dot{U}$ denotes dU/dz at $z=0$. In the other limit of a perfectly conducting ground ($\beta \ll 1$) the following expansion holds

$$D = D_c + \beta U^2 + \beta^2 U^2(\partial U/\partial D) + \cdots, \quad \text{(c)}$$

where D_c is the characteristic value obtained under the condition $\dot{U}(0)=0$. Since $\dot{U}$ and U in (a) and (c), respectively, are known from the normalization condition, it is possible to determine the first correction terms to the characteristic values in each case.

INTRODUCTION

IN the normal mode theory[1] of propagation of microwaves under non-standard conditions of atmospheric refraction, the characteristic values are derived under the assumption that the height-gain function U vanishes at the ground. This boundary condition is valid only when a parameter β, depending on the ground constants, is large, the so-called case of strong absorption, where

$$\beta_v = i(k_1^2 - k^2)^{\frac{1}{2}}(qk_0^2)^{-\frac{1}{2}}/\epsilon_1 \simeq i(k_0/q)^{\frac{1}{2}}(\epsilon_1 - 1)^{\frac{1}{2}}/\epsilon_1; \quad \beta_h = \beta_v \epsilon_1, \quad (1)$$

$$\epsilon_1 = \epsilon - i(\sigma/\omega); \quad k_0 = 2\pi/\lambda; \quad q = 2/a_e, \quad (2)$$

the subscripts v and h refer to vertical and horizontal polarization, respectively, and a_e denotes the effective radius of the earth. In Figs. 1 and 2 values of β_v^{-1} and β_h^{-1} are given for sea water and "average soil." It is seen that β_h^{-1} is always small, but that β_v^{-1} becomes large at the longer wave-lengths. The purpose of this investigation is primarily to ascertain the change in the characteristic values of the normal modes resulting from the finiteness of β. Such a correction, although small, is of importance under conditions of strong trapping and vertical polarization, where in the first approximation the theoretical horizontal decrement, which is proportional to the imaginary part of D, is extremely small. It turns out that the first correction term can be determined, without additional computation, from the known properties of the height-gain function.

FORMAL SOLUTION

The formal solution for the Hertzian potential Ψ resulting from a dipole radiating in an atmosphere under non-standard refraction conditions can be obtained along the lines followed[2] in the case of standard atmospheric refraction. The wave equation

$$\nabla^2\Psi + k_0^2\mu^2\Psi, \quad k_0 = 2\pi/\lambda, \quad (3)$$

is synthesized from elementary solutions of the form

$$\psi = e^{i\omega t} J_0(kr) F(h) G(k), \quad (4)$$

where, on replacing μ by the modified index of refraction N, F must satisfy the differential equations

$$d^2F/dh^2 + (k_0^2N^2 - k^2)F = 0, \quad h > 0, \quad (5)$$

$$d^2F/dh^2 + (k_1^2N^2 - k^2)F = 0, \quad h < 0, \quad k_1^2 = k_0^2\epsilon_1, \quad \epsilon_1 = \epsilon - i\sigma/\omega. \quad (6)$$

* This paper is based on work done for the Naval Research Laboratory under contract with the Navy's Office of Research and Inventions. Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University

[1] Radiation Laboratory Report 680 (1945) by W. H. Furry; C. L. Pekeris, J. Acous. Soc. Am. **18**, 304 (1946); Phys. Rev. **70**, 518 (1946); J. App. Phys. **17**, 1108 (1946).

[2] C. L. Pekeris, J. App. Phys. **18**, 667 (1947).

Let $M(h)$ and $N(h)$ denote the solutions of (5) which at great heights, h, represent downcoming and outgoing waves, respectively, and let $P(h)$ denote the solution of (6) which does not become infinite at great depth. If h_1 denote the height of the transmitter, then the appropriate solutions are

$$F_1(h)=C(k)N(h), \qquad h>h_1, \tag{7}$$

$$F_2(h)=A(k)M(h)+B(k)N(h), \qquad 0<h<h_1, \tag{8}$$

$$F_3(h)=E(k)P(h), \qquad h<0. \tag{9}$$

The functions A, B, C, and E of k are determined from the boundary conditions:

$$F_1=F_2; \quad F_2'-F_1'=2k, \quad \text{at } h=h_1, \tag{10}$$

$$\gamma F_2=F_3, \quad F_2'=F_3',$$

$$(\gamma_v=1/\epsilon_1, \quad \gamma_h=1), \quad \text{at } h=0, \tag{11}$$

where the prime denotes differentiation with respect to h. It is found that $P(h)$ enters the expression for the potential in the atmosphere only through the combination $P'(0)/P(0)$, for which one can write with very high approximation[3] $i(k_1^2-k^2)^{\frac{1}{2}}$. One thus obtains

$$\Psi_1=-(2e^{i\omega t}/b)\int_0^\infty J_0(kr)kdkN(h)$$

$$\times[M(h_1)+(B/A)N(h_1)], \quad h>h_1, \tag{12}$$

$$\Psi_2=-(2e^{i\omega t}/b)\int_0^\infty J_0(kr)kdkN(h_1)$$

$$\times[M(h)+(B/A)N(h)], \quad 0<h<h_1, \tag{13}$$

$$b\equiv MN'-NM'=\text{constant};$$

$$(B/A)=-[\beta M(0)-HM'(0)]/$$

$$[\beta N(0)-HN'(0)], \tag{14}$$

where β is defined in (1), and $H(=k_0^{-\frac{1}{3}}q^{-\frac{1}{3}})$ is the natural unit of height. In terms of $z=h/H$ we have

$$B/A=-[\beta M(0)-\dot{M}(0)]/Q, \tag{15}$$

$$Q\equiv[\beta N(0)-\dot{N}(0)]; \quad \dot{M}\equiv dM/dz. \tag{16}$$

The normal mode solution is obtained by evaluating the integrals (12) and (13) in terms of the residues[1] at the zeros k_m in the k-plane of Q:

$$\Psi=-(i\pi e^{i\omega t}/H)\sum_1^\infty H_0^{(2)}(k_m r)U_m(z_1)U_m(z), \tag{17}$$

where $U_m(z)[=N_m(z)]$ is normalized by the condition

$$\operatorname*{Lim}_{I(k)\to 0^-}-\int_0^\infty U_m^2(z)dz=1, \tag{18}$$

and must satisfy the boundary condition at the ground

$$\beta U_m(0)-\dot{U}_m(0)=0. \tag{19}$$

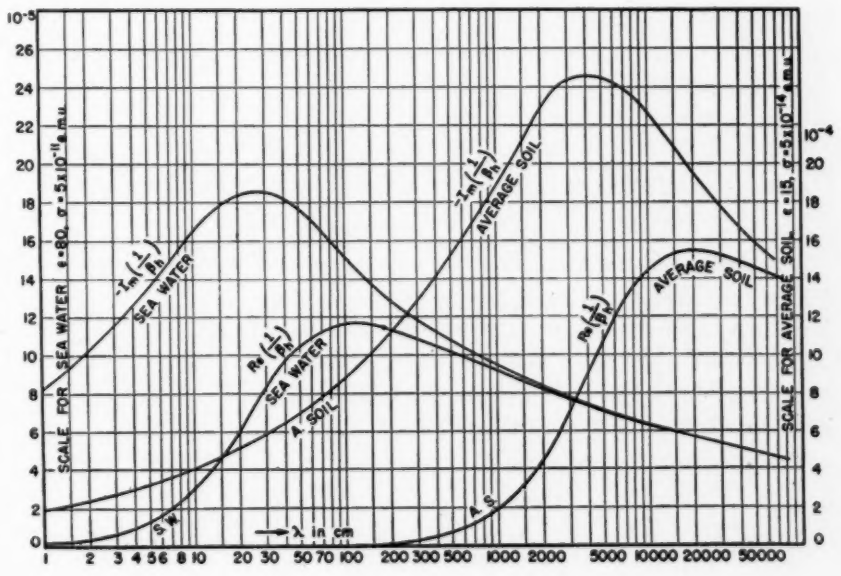

SCALE FOR SEA WATER ε=80, σ=5x10⁻¹¹ e.m.u.
SCALE FOR AVERAGE SOIL ε=15, σ=5x10⁻¹⁴ e.m.u.
-Im(1/βh) SEA WATER
Re(1/βh) SEA WATER
-Im(1/βh) AVERAGE SOIL
Re(1/βh) AVERAGE SOIL
A. SOIL
A. S.
S.W.
λ in cm
1 2 3 4 5 6 8 10 20 30 50 70 100 200 300 500 1000 2000 5000 10000 20000 50000

FIG. 1. Values of $(1/\beta_h)$ for "average soil":
$\epsilon=15$, $\sigma=5\times10^{-14}$ e.m.u.
sea water:
$\epsilon=80$, $\sigma=5\times10^{-11}$ e.m.u.

[3] See reference 2, p. 670.

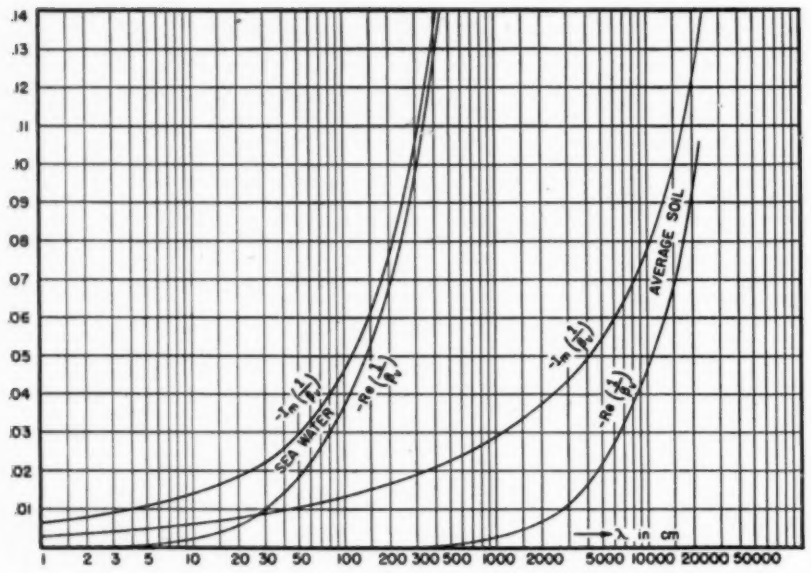


FIG. 2. Values of $(1/\beta_v)$ for "average soil":
$\epsilon=15$, $\sigma=5\times10^{-14}$ e.m.u.
sea water:
$\epsilon=80$, $\sigma=5\times10^{-11}$ e.m.u.

$U_m(z)$ is a solution of (5), which by the substitution

$$N^2(h)=1+(q/k_0)^{\frac{1}{3}}Y(z);$$
$$k^2=k_m^2=k_0^2[1-(q/k_0)^{\frac{1}{3}}D_m], \qquad (20)$$

is transformed into

$$d^2U_m(z)/dz^2+[D_m+Y(z)]U_m(z)=0. \qquad (21)$$

DEVELOPMENT OF D IN POWERS OF β

If, for a given $Y(z)$, the N-type solution of (21) is integrated from infinity for various assumed values of D, and—after normalization by (18)—the quantities $U_m(0, D)$ and $\dot{U}_m(0, D)$ are obtained as functions of D, the characteristic value D_m will be the one for which condition (19) is satisfied. We shall, in the sequel, denote by $U(D)$ and $\dot{U}(D)$ the quantities $U_m(0, D)$ and $\dot{U}_m(0, D)$, respectively. When β is large, the boundary condition (19) reduces to

$$U(D)=0, \quad D=D_a, \qquad (22)$$

where the subscript a refers to strong absorption. Hitherto, characteristic values for non-standard propagation have been determined[4] from condition (22). In order to ascertain the effect on D of the finiteness of β, we put

$$D-D_a=(D_1/\beta)+(D_2/\beta^2)+\cdots, \qquad (23)$$
$$U(D)=U(D_a)+(D-D_a)(\partial U/\partial D)+\cdots, \qquad (24)$$
$$\dot{U}(D)=\dot{U}(D_a)+(D-D_a)(\partial \dot{U}/\partial D)+\cdots, \qquad (25)$$

substitute into (19), and equate the coefficients of powers of β to zero. One thus finds that

$$D_1=\dot{U}/(\partial U/\partial D);$$
$$\frac{\partial U}{\partial D}D_2=-(1/2)D_1^2\frac{\partial^2 U}{\partial D^2}+D_1\frac{\partial \dot{U}}{\partial D}. \qquad (26)$$

D_1 can be evaluated in terms of known quantities by using a standard procedure in boundary value problems. Consider two solutions, U_1 and U_2, of (21) obtained for two assumed values, D_1 and D_2 of D_m. If we multiply the two equations

$$d^2U_1/dz^2+[D_1+Y(z)]U_1=0, \qquad (27)$$
$$d^2U_2/dz^2+[D_2+Y(z)]U_2=0, \qquad (28)$$

by U_2 and U_1, respectively, subtract and then integrate the result, we arrive at

$$(U_1\dot{U}_2-U_2\dot{U}_1)_{z=0}=(D_2-D_1)\int_0^\infty U_1U_2dz. \qquad (29)$$

By now writing

$$U_2=U_1+(D_2-D_1)(\partial U_1/\partial D)+\cdots;$$
$$\dot{U}_2=\dot{U}_1+(D_2-D_1)(\partial \dot{U}_1/\partial D)+\cdots, \qquad (30)$$

and passing to the limit $D_2\to D_1$, we get

$$U(\partial \dot{U}/\partial D)-\dot{U}(\partial U/\partial D)=\int_0^\infty U^2dz=1. \qquad (31)$$

Upon differentiating (31) with respect to D, we

[4] C. L. Pekeris, Proc. I.R.E. **35**, 453 (1947); C. L. Pekeris and S. Ament, Phil. Mag. (1948).

obtain

$$U(\partial^2\dot{U}/\partial D^2)-\dot{U}(\partial^2 U/\partial D^2)=0. \quad (32)$$

Since $U(D_a)$ vanishes, it follows from (31) and (32) that at D_a

$$(\partial U/\partial D)=-1/\dot{U}; \quad \partial^2 U/\partial D^2=0. \quad (33)$$

When substituted in (26) and (23), this yields

$$D_1=-\dot{U}^2; \quad D_2=\dot{U}^3(\partial\dot{U}/\partial D), \quad (34)$$

$$D=D_a-(\dot{U}^2/\beta)+\dot{U}^3(\partial\dot{U}/\partial D)/\beta^2+\cdots. \quad (35)$$

Now $\dot{U}$ is required in determining the height-gain function $U(z)$, so that the first correction term in (35) does not require any additional computation.

For the case of a perfect conductor ($\beta\ll 1$), the boundary condition (19) reduces to

$$\dot{U}(D)=0, \quad D=D_c. \quad (36)$$

One may therefore expand D into the form

$$D=D_c+\beta D_1+\beta^2 D_2+\cdots, \quad (37)$$

where it is found that

$$D_1=U/(\partial\dot{U}/\partial D);$$

$$D_2\frac{\partial\dot{U}}{\partial D}=D_1\frac{\partial U}{\partial D}-\frac{1}{2}D_1^2\frac{\partial^2\dot{U}}{\partial D^2}. \quad (38)$$

Using now (31) and (32), we find that at D_c

$$U(\partial\dot{U}/\partial D)=1; \quad \partial^2\dot{U}/\partial D^2=0, \quad (39)$$

and, therefore,

$$D_1=U^2; \quad D_2=U^3(\partial U/\partial D), \quad (40)$$

$$D=D_c+\beta U^2+\beta^2 U^3(\partial U/\partial D)+\cdots. \quad (41)$$

Here again U is known if the normalized height-gain function is determined.

Errata: The Electrostatic Field of a Point Charge inside a Cylinder, in Connection with Wave-Guide Theory

[J. App. Phys. 18, 562 (1947)]

C. J. Bouwkamp and N. G. de Bruijn
Eindhoven, Netherlands

THE last part of the paper, relating to scalar wave problems, needs some correction. It should be borne in mind that, in boundary-value problems of the second kind ($\partial V/\partial\rho=0$ at $\rho=a$), the function $V=\exp(ikz)$ is always a possible solution of the wave equation satisfying the boundary condition mentioned. In other words, for acoustic waves the cylinder does not behave as a high pass filter; the lowest cut-off frequency is zero.

For instance, the field of an axial point source ($u_0=e^{ikR}/R$) inside the rigid cylinder is given by

$$u=\frac{2i}{ka^2}e^{ik|z|}+\frac{2}{a}\sum_{n=1}^{\infty}\frac{y_0(\rho x_n/a)}{y_0^2(x_n)}\frac{e^{-|z|/a(x_n^2-k^2a^2)^{\frac{1}{2}}}}{(x_n^2-k^2a^2)^{\frac{1}{2}}}$$

where the first term originates from x_0, x_n being the set of non-negative zeros of $J_1(x)$. The same function u represents z-component of the magnetic potential of anaxial-magnetic dipole inside a perfectly conducting cylinder. In that case, however, the first term does *not* contribute to the electromagnetic field itself.

The following misprints have been noticed.

Eq. (4): $z^{-i\lambda z}$, read $e^{-i\lambda z}$;
Eq. (c'): $(-1)\nu$, read $(-1)^\nu$;
First line following Eq. (27): $x=0$, read $z=0$;
Eq. (35): E^z, read E_z;
Eq. (42): Y, read J;
Page 575: u_1, u_2, read U_1, U_2;
Fourth line following (C) of "Case 1": κ, read x.

A Study on Anaeroid Capsules*

EUGENE W. PIKE** AND NORMAN E. GIBBS***
The R. W. Cramer Company, Inc., Centerbrook, Connecticut
(Received September 2, 1947)

The analysis of anaeroid capsule performance by the expansion of the observed pressure-deflection curve in Gram summation-orthogonal polynomials is illustrated on a series of fifteen runs on five typical capsules.

All of the deflection curves examined contained significant cubic and quartic terms while a few showed fifth-order terms. These high order terms are very unstable in magnitude and sign for successive deflections of the same capsule, and this accounts for the very complex hysteresis curves which have been observed.

The relationship of these high order terms in the deflection characteristic to the accuracy and cost of precise anaeroid barometers is brought out, and it is suggested that barometric systems free of elastic redundancies must be devised if accurate barometers for radiosondes and similar instruments are to be mass produced.

I. INTRODUCTION

THE rapidly increasing use of radiosondes and similar instruments in meteorology has created a demand for precise, low cost, mass-produced anaeroid barometers. This paper is a preliminary survey of the difficulties in the way of such a development, and a presentation of some useful methods for the study of capsule performance (and indeed, for the study of deflection instruments in general).

II

The anaeroid capsules used in this study consisted of two three-cell beryllium copper capsules, manufactured by the Kollsman Instrument Division of the Square D Company, and two two-cell Elinvar capsules, manufactured by the Mercoid Corporation. (The beryllium copper capsule "A" had a bimetal linkage for temperature compensation attached to it.) In addition, a single run was made on a captured German brass capsule, marked only "KST."

The deflection was measured with a Carson electronic micrometer,† a precise screw micrometer driven by a slow speed motor which is in turn controlled by an electronic circuit, so that the micrometer hunts around a position which maintains a contact resistance of about 50 megohms between the capsule and moving element of the micrometer.

This micrometer was mounted in a bell jar in which the pressure could be very accurately controlled by a vacuum pump and an arrangement of needle valves.

The pressure was read on a 7″ Wallace and Tiernan precision anaeroid manometer, which was, in turn, calibrated each month against a standard mercurial manometer at the Evans Signal Laboratory. The same exact schedule of pressure and time was used in calibrating and using the manometer.

It was not possible, in the time available, to arrange for thermostatting of the bell jar, so that there was some distortion of the micrometer from the heat of the motor, as well as change in the elastic modulus of the beryllium copper capsules. These changes were slow and regular, however, and do not affect the conclusions reached in the body of this paper.

III

The deflection of the capsule was measured at 50-millibar†† pressure intervals from 1000 millibars to 50 millibars at the rate of one reading every three minutes exactly.

* The observations reported herein were obtained as part of Contract W-28-003-sc-1531 between The R. W. Cramer Co., Inc. and the Signal Corps Laboratories. Interrupted by the end of the war, these observations were reduced and prepared for publication by the authors privately. Publication is by permission of the Signal Corps.

** Now with McFarlan, Groth, and Pike, P.O. Box 81, Grand Central Station, New York 17, New York.

*** Now with Reaction Motors Inc., Dover, New Jersey.

† Available through Instrument Specialties Company, Little Falls, New Jersey.

†† The millibar, a meteorological unit of pressure, is 1000 dynes per cm². 1000 mb is very closely one atmosphere.

The twenty points of each deflection curve were then expanded in the gram summation-orthogonal polynomials[1] until the residuals were random, i.e., showed no ordered structure.†††

The amplitudes of the successive polynomials, and the root-mean-square "$\bar{\Delta}$" of the residuals, in millibars, are given in the left-hand six columns of Table I.

While the linear and quadratic terms, a_1 and a_2, change very little from run to run, the higher order terms are relatively very unstable. The amplitude of the cubic term even changes sign in the case of Elinvar *B*. It will be noticed also that the mean amplitude of the residuals, in millibars, is quite constant (with the exception of Elinvar *A*) in spite of the large changes from capsule to capsule in all other parameters.

In a few cases, polynomials were fitted to alternate points only, ten in all, as well as to all twenty of the observed points. The sums of the squares of the residuals (in units of ten micro inches) for the corresponding ten-point and twenty-point expansions appear in the last two columns on the right of Table I. The noteworthy point is the irregularity of the increase in the residuals for the ten point expansions.

Corresponding residuals for rising and falling pressures, and corresponding residuals for successive runs in the same direction, were tested for possible correlation. The results are shown in Table II. None of the correlations are significant except the tracking correlations for BeCu"A." Examination of the original data shows that this is the only case for which the actual micrometer readings were closely the same at corresponding pressures for the two correlated runs.

The conclusion from the correlation tests and the constancy of the mean residuals is that the random residuals are primarily the result of short-period random errors in the micrometer screw, and that the deviation of the actual capsule deflection from a smooth curve is not greater than perhaps 0.05 millibars.

[1] R. A. Fisher and F. Yates, *Statistical Tables* (Oliver and Boyd, London, 1943), second edition, p. 20 and Table XXIII.

††† A number of tests for the randomness of the residuals were used: the serial correlation test (R. L. Anderson: Ann. Math. Stat. **13**, 1 (1942)); the tests for runs up and down, (F. Swed and C. Eisenhart: Ann. Math. Stat. **14**, 66 (1943); W. A. Wallis, and G. H. Moore, J. Am. Stat. Soc. **38**, 153 (1943)); the test for runs on either side of median (F. Mosteller; Ann. Math. Stat. **12**, 228 (1941)).

TABLE I. The amplitudes a_n of the successive gram polynomials, and the root-mean-square "$\bar{\Delta}$" of the residuals, derived by expanding the pressure-deflection measurements for 15 runs on 5 typical anaeroid capsules in gram summation-orthogonal polynomials.‡

Capsule‡‡		a_1	a_2	a_3	a_4	a_5	"$\bar{\Delta}$" (mb)	Residuals‡ 20 Pt.	10 Pt.
"KST"	*F*	6756	−49.3	− 20.9	—	—	0.32	407	444
BeCu "A"									
9/23	*F*	9447	768	25.6	− 7.4	—	0.28	498	875
	R	9449	751	17.7	−12.4	—	0.23		
9/24	*F*	9457	778	20.9	− 8.7	—	0.25	504	607
	R	9453	756	14.9	−13.6	−2.4	0.17		
BeCu "B"									
9/21	*F*	9418	764	−101.1	−47.0	—	0.21		
	R	9426	754	−111.4	−39.1	—	0.24		
	F	9363	767	−105.1	−40.2	—	0.22	301	991
	R	9411	747	−106.9	−43.4	—	0.18	172	608
Elinvar "A"									
9/21	*F*	8447	875	8	−31	—	0.46	965	1163
	R	8457	854	5	−29	−5	0.57		
Elinvar "B"									
9/22	*F*	8542	891	4	−31	—	0.29		
	R	8542	857	− 9	−26	—	0.25	230	230
9/23	*F*	8539	891	—	−22	−5	0.26		
	R	8542	851	− 13	−28	−3	0.18		

‡ The two columns on the right give the sum of the squared residuals for polynomials fitted to all 20 observed points, and to only 10 alternate points, respectively.

‡‡ The dates in the "capsule" column refer to the day the run was made, the "*F*" and "*R*" to falling and rising pressure, respectively.

IV

The relationship of the data in Table I to the accuracy and cost of precise pressure measurements is immediately plain, when the train of inference involved in a pressure determination is examined. From the observed deflection of a capsule, the pressure is inferred by reference to a calibration curve, drawn through a series of points observed during a previous deflection. There are five sources of error in this procedure:

(a) The calibrating instruments are not perfectly accurate.
(b) The calibration curve (either a graph or a fitted polynomial) does not represent the true deflection curve exactly, especially between calibration points (Interpolation error).
(c) The capsule does not repeat its deflection exactly on the second excursion (Tracking error).
(d) The deflection of the capsule is not read exactly.
(e) The calibration curve is not read exactly.

TABLE II. The coefficient of correlation between the residuals for falling and rising pressure ("Hysteresis"), and between the residuals for successive runs in the same direction of pressure change ("Tracking").

Capsule	Hysteresis 9/23	Hysteresis 9/24	Tracking *F*	Tracking *R*
Elinvar "B"	0.22	0.24	0.23	0.05
BeCu "A"	0.29	0.36	0.86	0.76

Up to the present time, almost all the effort in capsule development has been placed upon reduction of error (c). Table I shows that, aside from the obviously temperature inspired change in a_1 for the beryllium copper capsule "B," the tracking error is almost wholly caused by changes in the higher order terms in the expansion. If these terms could be eliminated, the tracking error would be markedly reduced.

A calibration curve which is known to be exactly parabolic can be fitted with adequate accuracy to eight calibration points, whereas at least 20 calibration points are required to determine a calibration curve which may have a quintic term. Since about 85 percent of the cost of a precise barometer of the mass-produced type used in radiosondes is the cost of calibrating the instrument and preparing the calibration curve, either by expansion in polynomials or by accurate graphical methods, it seems obvious that, in a barometer intended to indicate pressure with a mean error of half a millibar or less, a very expensive mechanical structure can be tolerated if it eliminates the higher order terms in the deflection curve, and so reduces the cost of calibration by half to two-thirds.

The exact cause of the unstable, high order terms in the deflection curves is not known. Two mechanisms have been suggested: elastic redundancy leading to over-center movements, "oil-canning," (such movements are clearly observable in some very sensitive capsules), and bending of the axis of the capsules with change in pressure.

Whatever the cause, it seems likely that a mechanically and elastically simple structure is most likely to give simple deflection curves. Previous attempts to achieve this simplicity by the use of thin capsules and external flat strip restoring springs have not been successful, but there is no reason to regard such a structure as elastically more simple than a capsule. The writers would suggest a capsule built like an accordion, with two rigid end pieces of constant effective area, joined by a very soft bellows, and restrained by a compound coil spring of the type used in precision spring balances.

It is a pleasure to acknowledge the aid and encouragement of the staff of the Meteorological Branch, Evans Signal Laboratory, and the patience of Miss F. M. Wright who made most of the measurements.

BIBLIOGRAPHY

The following papers have been useful:

"Investigation of the Diaphragm-Type Pressure Cell," N.A.C.A. Technical Report 388.

W. A. Wildhack and V. H. Goerke, "The Limiting Useful Deflections of Corrugated Metal Diaphragms," N.A.C.A. Technical Note 876.

"Corrugated Metal Diaphragms for Aircraft Pressure Measuring Instruments," N.A.C.A. Technical Report 738.

S. Dushman, "Creep of Metals," J. App. Phys. **15**, 108 (1944).

Paul G. Exline, "Pressure Responsive Elements," Trans. A.S.M.E. **60**, 625 (1938).

Calendar of Meetings

January, 1948

29–31 American Physical Society, New York, New York

February

2–5 American Society of Heating and Ventilating Engineers, New York, New York
6 American X-Ray and Electron Diffraction Society (Michigan Chapter), Ann Arbor, Michigan
6–7 American Philosophical Society, Philadelphia, Pennsylvania
13–14 American Geophysical Union, University, University of California, Berkeley, California
15–19 American Institute of Mining and Metallurgical Engineers, New York, New York
23–26 Technical Association of the Pulp and Paper Industry, New York, New York
28 American Mathematical Society, New York, New York and Chicago, Illinois

March

1–5 American Society for Testing Materials, Washington, D. C.
2–3 Inter-Society Color Council, New York, New York
4–6 Optical Society of America, New York, New York
22–24 American Chemical Society and the Institute of Technology of the University of Minnesota, Symposium "Modern Instrumental Methods of Analysis," Minneapolis, Minnesota
22–25 Institute of Radio Engineers, New York, New York
26 Institute of Aeronautical Sciences (Flight Propulsion Meeting), Cleveland, Ohio

April

1–3 American Institute of Electrical Engineers (Great Lakes District Meeting), Des Moines, Iowa
7–9 American Society of Civil Engineers (Spring Meeting), Pittsburgh, Pennsylvania
8–10 Geological Society of America (Cordillera Section), Pasadena, California
16–17 American Mathematical Society, New York, New York and Ann Arbor, Michigan
17 American Mathematical Society, Berkeley, California
19–23 American Chemical Society, Chicago, Illinois
21–23 American Geophysical Union, Washington, D. C.
22–24 Acoustical Society of America, Washington, D. C.
22–24 American Philosophical Society, Philadelphia, Pa.
25–30 American Ceramic Society, Inc., Chicago, Illinois
26–28 National Academy of Sciences, Washington, D. C.
28–30 American Institute of Electrical Engineers (Northeastern District Meeting), New Haven, Connecticut
30–May 1 Canadian Institute of Radio Engineers, Toronto, Ontario, Canada

Letters to the Editor

Remarks on Slow Waves in Cylindrical Guides

A. A. Oliner
Microwave Research Institute, Polytechnic Institute of Brooklyn, Brooklyn, New York
October 1, 1947

SEVERAL articles, one in 1944[1] and two more recently,[2] have appeared on the question of concentric dielectrics in circular wave guides, each treating it from a different point of view. The problem considered by the more recent articles is the utilization of the TM_{01} mode in a circular wave guide with two coaxial dielectrics to obtain wave propagation with a phase velocity below that of light velocity. A main objective of each of these articles is to obtain a method whereby one is able to determine the necessary values of the dielectric constants and radii involved, when the phase velocity is given some pre-assigned value. The earlier article considers the problem more generally. All of the authors approach the problem by solving for the fields and then applying the appropriate boundary conditions.

The purpose of this letter is twofold; (1) to indicate that an alternative and simpler method may be used to derive the necessary relations, and (2) to acquaint the reader with the existence of certain curves whose use greatly eases the calculations.

The properties of an ideal wave guide are determined once the cut-off frequency is determined. The cut-off frequency for a given mode corresponds to resonance for waves propagating transversely in the cross section. Therefore, to find the cut-off frequency of the guide, it is necessary only to determine the resonant frequency of the two-dimensional problem defined by the guide cross section. The condition for resonance that we apply is that, at any radius, the admittances looking in opposite radial directions should be equal and opposite.

Since the above-mentioned articles do not agree as to notation, it is in place to specify the notation used here. The inner and outer regions shall be regions 1 and 2, respectively (see Fig. 1). Dielectric constant ϵ_2 is greater than ϵ_1. k_c shall represent the cut-off wave number, and β, the propagation constant in the axial direction. We then have:

$$k_{c1}=(k^2\epsilon_1-\beta^2)^{\frac{1}{2}}, \quad k_{c2}=(k^2\epsilon_2-\beta^2)^{\frac{1}{2}}, \tag{1}$$

where $k=\omega/c$ and $\beta=\omega/v_p$, v_p being the phase velocity.

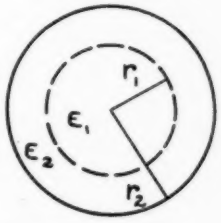


FIG. 1.

A radial guide is the appropriate one to use in the cross section since we have propagation down a circularly cylindrical wave guide. The expression for the admittance $Y(r_1)$ at radius r_1 in terms of the admittance $Y(r_2)$ at r_2, for the lowest E type mode in a radial guide can be written in analogy with the uniform line case as:[3]

$$Y(r_1)=Y_0(r_1)\frac{j+\zeta(x,y)ct(x,y)\,Y(r_2)/Y_0(r_2)}{Ct(x,y)+j\zeta(x,y)\,Y(r_2)/Y_0(r_2)}, \tag{2}$$

where

$$\zeta(x,y)=\frac{J_0(x)N_0(y)-N_0(x)J_0(y)}{J_1(x)N_1(y)-N_1(x)J_1(y)},$$

$$ct(x,y)=\frac{J_1(x)N_0(y)-N_1(x)J_0(y)}{J_0(x)N_0(y)-N_0(x)J_0(y)}, \tag{3}$$

$$Ct(x,y)=\frac{J_1(y)N_0(x)-N_1(y)J_0(x)}{J_1(x)N_1(y)-N_1(x)J_1(y)},$$

$x=k_cr_1$, $y=k_cr_2$, and Y_0 is the characteristic admittance of the guide which is a function of the radius. It may be seen that if a short circuit exists at r_2, then $Y(r_2)=\infty$, and $Y(r_1)$ reduces to:

$$\bar{Y}(r_1)=-jY_0(r_1)ct(x,y). \tag{4}$$

Similarly, we find that for any finite $Y(r_2=0)$:

$$\bar{\bar{Y}}(r_1)=jY_0(r_1)\frac{J_1(x)}{J_0(x)}. \tag{5}$$

The condition for resonance is satisfied if, at radius $r=r_1$, we set $\bar{Y}+\bar{\bar{Y}}=0$.
We then obtain:

$$Y_{01}(r_1)\frac{J_1(k_{c1}r_1)}{J_0(k_{c1}r_1)}=Y_{02}(r_1)ct(x,y), \tag{6}$$

where $x=k_{c2}r_1$ and $y=k_{c2}r_2$. For E type modes the ratio of the characteristic admittances is given by

$$\frac{Y_{01}(r_1)}{Y_{02}(r_1)}=\frac{\epsilon_1k_{c2}}{\epsilon_2k_{c1}}. \tag{7}$$

We may note that k_c is the propagation constant in the transverse region, although β is the propagation constant in the axial direction. Substituting for the characteristic admittances, we obtain finally:

$$\frac{\epsilon_1k_{c2}}{\epsilon_2k_{c1}}\frac{J_1(k_{c1}r_1)}{J_0(k_{c1}r_1)}=ct(x,y), \tag{8}$$

where $x=k_{c2}r_1$ and $y=k_{c2}r_2$.

It may be seen that Eq. (8) is identical, except for notation, with Eq. (8) of the article by Bruck and Wicher. These authors state that this expression is the relation between the field velocity, tube radii, and dielectric constant, from which they obtain their results. They further state that this relation "can be solved by a combination of patience with a variety of numerical procedures." However, there exist curves of the combination of Bessel functions represented by $ct(x,y)$. (See Fig. 2.) The values of $ct(x,y)$ are plotted against $y-x$, with y/x as a parameter. The results desired may be obtained quite rapidly by the use of these ct curves. The method involved here is essentially the same as that used by these authors; the curves simply serve to facilitate the calculations.

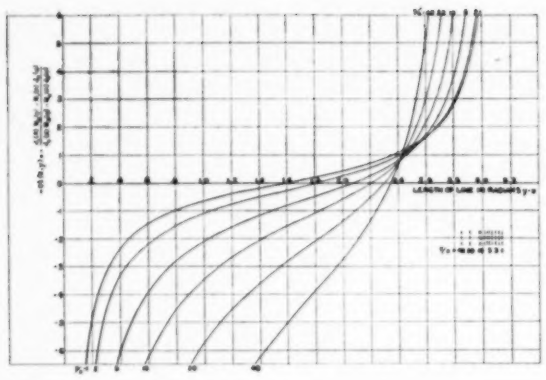


FIG. 2.

As an example, one of the points considered by the above authors was checked using the *ct* curves. The phase velocity chosen was $v_p=0.2c$, and the smaller and larger radii were 0.05λ and 0.10λ, respectively. The required ratio of dielectric constants must be determined. The value $\epsilon_2/\epsilon_1=57.3$ was obtained by means of the *ct* curves. Bruck and Wicher state that $\epsilon_2/\epsilon_1=57.25$.

In the article by Frankel, the final relations are obtained in somewhat different form. However, after appropriate rearranging of terms, Eq. (28) of Frankel's article may be seen to be identical, again except for notation, with Eq. (8) above. In his "typical example," Frankel specifies the values of all the necessary parameters except the larger radius, which must be solved for. His result, obtained graphically, is 3.8 millimeters. The result obtained by means of the *ct* curves is 3.77 millimeters. Frankel states further that "a more careful calculation using tables rather than curves" shows that the result should be 3.6 millimeters. The writer believes that this statement is in error as the use of tables with Eq. (8) above yields a value of 3.773 millimeters.

[1] L. Pincherle, "Electromagnetic waves in metal tubes filled longitudinally with two dielectrics," Phys. Rev. **66**, 118 (1944).
[2] Sidney Frankel, "TM_{01} mode in circular wave guides with two coaxial dielectrics," J. App. Phys. **18**, 650 (1947); G. G. Bruck and E. R. Wicher, "Slow transverse magnetic waves in cylindrical guides," J. App. Phys. **18**, 766 (1947).
[3] A brief derivation may be found in Vol. 8, Chap. 8, of the Radiation Laboratory Series (in print).

Specimen Contamination in Electron Microscopes

JOHN H. L. WATSON
Research Institute of the Henry Ford Hospital, Detroit 2, Michigan
November 7, 1947

IN a recent letter to the editor of the Journal of Applied Physics, Cosslett[1] has advanced the hypothesis that specimen growth resulting from contamination is the result of the deposition of charged metal and other particles ejected from the supporting grid by the electron beam. We[2] have suggested the hypothesis that the growth is caused by the deposition as solids of hydrocarbon vapors polymerized under the bombardment and originating from locations inside the electron microscope including the surfaces of the walls, the specimen holder, the supporting grid, the specimen itself, or perhaps the oil diffusion pump. Both parties agree that the contaminant is more efficiently deposited as the intensity is raised and as the time interval increases. Our explanation would be contained presumably in that portion of Cosslett's which is implied in the phrase "other particles ejected."

It is well known that temperatures do rise to considerable values in the instrument. We have melted certain pieces of titanium dioxide which would indicate that it is possible to attain local temperatures of the order of 1600°C. We have not observed fusion of the supporting grids even with a biassed gun, probably because of good thermal contacts. However, temperatures can rise to a degree commensurate with an explanation such as Cosslett advances.

If his arguments are correct, they must not only explain satisfactorily the observations given in reference 2 but they must also satisfy the arguments given below. It is to be pointed out that all of the observations cited by Cosslett can find an explanation under our assumptions of hydrocarbon deposition.

Although contamination is *less* over a film at locations far from the grid, there is still contamination (film thickening) wherever the beam strikes the surface to a greater or less extent depending upon conditions. For example, Fig. 8 of reference 2 is of a clean film which is supporting no specimen but which shows definite film thickening. This phenomenon can be readily demonstrated and is often a cause for lowered contrast in electron micrographs. Likewise, although contamination is much enhanced near the grid, it would be risky to say that it could not occur far from the grid even upon filmless specimens.

Deposits similar to these are found on bombarded surfaces in general throughout electronic systems. Although these will form in time, regardless of precautions, we have found unmistakable decreases in them in the electron microscope when regular and thorough cleaning of the inside of the instrument was carried out, using benzene or other solvents with fine abrasives.

Hillier[3] has recorded several facts concerning the phenomenon which seem to run contra to the hypothesis of Cosslett. He finds that the contaminant layer around a particle of *carbon* yields less contrast in an electron micrograph than does the particle itself. For this to be so, the contaminating layers must be either more loosely agglomerated or of equal or less physical density. A combination of these two explanations or either singly, could account for the lighter areas surrounding Cosslett's zinc oxide crystals, but with *carbon* as the contaminated object, it is hardly likely that metallic coating of the thickness involved would show up as anything but a denser region surrounding the particle. These contaminations do not seem to be deposited as loose agglomerations as witnessed by the fact that the film-thickening layers are structureless in electron micrographs, regardless of their thickness.

Therefore, the lower opacity to electrons of the contamination would appear to be due to less physical density rather than to looseness of structure.

Again, Hillier with an electron probe of less than the usual angle has succeeded in depositing contaminant upon a very localized area of a film or specimen far from the grid. The probe was not striking the grid wires.

Hillier has observed further that with sufficient time the space inside the microscope clears and deposition ceases if the beam is not striking the grid. We have not been able to demonstrate this discontinuity ourselves but, since conditions governing the phenomenon are not all known or understood, it is not easy to duplicate them in order to demonstrate any particular circumstance. Such an occurrence would be indicative of either a cluster-ion or a free radical mechanism for the polymerization. In the original paper it was not intended to emphasize the cluster-ion hypothesis over that of the free radical, but the seeming dependence on conductivity tended to favor the former. If the dependence on conductivity is not real, either mechanism could account for polymerization and condensation from the gas phase, but the latter would be more probable.

The effect takes place under suitable circumstances to some extent with all materials but it is interesting that Cosslett is able to order substances according to the magnitude of the effect upon them. In our case the phenomenon was found to be so variable and conditions so difficult to control or analyze that such conclusions could not be drawn except to say that, in general, carbon blacks seem to be more affected than other materials. This conclusion appears to be supported by Cosslett's micrographs which demonstrate somewhat smaller changes in about the same time intervals.

Undoubtedly, as mentioned in reference 2, contamination can come from the specimen itself, but for the carbons examined the magnitude of the measured effect was not a function of the percentage of volatile matter known to be in the blacks. As for the Shawinigan black which is made by thermal decomposition at high temperature, it is known that there is but an extremely minute amount of non-carbon material in it.

Figure 1 is offered as an example of the appearance of materials which are redeposited after heating and evaporation by the bombardment of the electron beam. This shows a specimen of tetra copper calcium oxychloride which has been illuminated at high intensity and has changed considerably. In addition to the changes which have taken place over the face of the crystal itself, and which are incidental to the present discussion, it will be seen in the regions near the crystal and reaching out from it that material ejected or evaporated from the original location has been redeposited on the film. It will be noted that this "contaminant" has a structure and a granular appearance quite unlike the contaminants under discussion in this letter. If the specimen contaminations are caused by ejected or sputtered metallic particles, it might be reasonable to assume that they should have an appearance in some way similar to that in Fig. 1, but they do not.

[1] V. E. Cosslett, J. App. Phys. **18**, 844 (1947).
[2] John H. L. Watson, J. App. Phys. **18**, 153 (1947).
[3] Dr. James Hillier of the Radio Corporation of America has kindly given permission to mention his unpublished work. His full observations on contamination are to appear shortly.

FIG. 1. Tetra copper calcium oxychloride evaporated under electron bombardment and redeposited over the supporting film (×50,000).

The Use of a Cavity Oscillator for a Kerr Electro-Optical Light Shutter*

A. C. LAPSLEY, L. B. SNODDY AND J. W. BEAMS
Rouss Physical Laboratory, University of Virginia, Charlottesville, Virginia
October 8, 1947

THE successful use of a single cavity oscillator for the acceleration of electrons to energies of over a half million volts[1,2] suggested that a modification of this method might be used for producing an ultra high speed electro-optical Kerr cell light shutter. The high frequency high potential which occurs across a small gap inside the cavity oscillator should produce a large double refraction in liquids which exhibit an appreciable Kerr effect and in which the lag in the Kerr effect is sufficiently small.

The apparatus used in these experiments is similar to that used by Hereford[3] except that the cavity was filled with carbon disulphide. The cylindrical re-entrant type cavity was made of brass and was designed to resonate in the 400-megacycle range when filled with carbon disulphide. Essentially, the cavity may be considered as two quarter-wave concentric line resonators placed back to back with a small spacing between the inner conductors of the two lines. When properly matched to a generator a

high frequency field occurs across the gap between these inner conductors. The dimensions of the cavity were as follows: Outer conductor diameter, 15.5 cm; length of cavity, 20.3 cm; inner conductor diameter, 2.54 cm; spacing between inner conductors adjustable from 1.3 to 5.1 cm. The generator was a Western Electric 7C22 twin triode oscillator which delivers a peak power of 500 kilowatts with pulse operation and can be tuned from 390 to 425 megacycles. The tube was pulsed for 6×10^{-6} sec. 60 times per sec. by a conventional pulsing circuit with an output voltage of 18 kv. Power was transmitted from the tube through a $1\frac{1}{4}''\times3''$ coaxial 50-ohm transmission line to a properly oriented coupling loop inside of the cavity. Cavity resonance was observed by the usual crystal detector arrangement.

Light from a magnesium spark in series with the high voltage line to the tube was plane polarized by a Nicol prism so oriented that the electric vector of the light made an angle of 45° with the electric field in the cavity gap. The light passed through small strain free windows on opposite sides of the cavity gap and then into a second Nicol prism crossed with respect to the first. When the cavity was not in resonance, no observable light passed the second Nicol but, as soon as resonance occurred, light was observed to pass the second Nicol, showing that the cavity was operating as a Kerr cell light shutter. Carbon disulphide was used in the cavity because it has a medium-sized Kerr constant, is a good insulator without large dielectric loss at 400 megacycles and is known[3] to have a lag in the Kerr effect of less than 10^{-8} sec. It is planned to test other liquids with the hope of finding a better one, as well as to study the Kerr effect itself at high frequencies.

* This work was supported by the Bureau of Ordnance, U. S. Navy, under Contract NOrd-7873.

[1] Bowen, Pulley and Gooden, Nature **157**, 840 (1946).
[2] Hereford, Dissertation, University of Virginia (1947), J. App. Phys. (in press).
[3] Beams and Lawrence, J. Frank. Inst. **206**, 169 (1928).

New Books

Los Rayos X y La Estructura Fina de los Cristales. Fundamentos Teoricos y Metodos Practicos. (X-Rays and the Fine Structure of Crystals. Theoretical Fundamentals and Practical Methods.)

By Julio Garrido (del Instituto National de Física y Quimica de Madrid) y Joaquin Orland, S.J. (Professor de Metalografia en el Instituto Católico de Artes e Industrias). Pp. 260, 21×27 cm. Editorial Dossat, S. A., Plaza de Santa Ana 9, Madrid, 1946.

This is one of a half-dozen post-war foreign books which have appeared on the general subject of the application of x-ray diffraction to the study of crystals and crystalline aggregates. This particular book is unique in confining its coverage to the application of x-ray diffraction to the determination of crystal structures. It is an authoritative treatment of the subject written by authors who are known for their contributions in the field. It is not only well written, but it is profusely illustrated by well chosen linecuts and photographs. Together with Bunn's excellent book, it emphasizes the crystallographer's rather than the pure physicist's approach, and it is, therefore, a most welcome contribution to crystallographic literature.

The book, which contains ten chapters, is subdivided into three parts: Part one is a discussion of general principles; part two contains a discussion of the determination of crystal structures, while part three is entitled, "Some results of the study of the fine structure of matter by means of x-rays."

Part one contains a two-chapter development of the origin and physics of x-rays. This is followed by a chapter on the fundamentals of crystallography. The rest of part one consists of a very long chapter on the diffraction of x-rays by crystals. The whole section is very well done, and the chapter on crystallography is particularly well treated.

Part two contains chapters on the various x-ray diffraction methods, on the sequence of experiment and argument in the determination of a crystal structure, and examples of structure determination. This part constitutes the heart of the book, and should prove extremely useful to the student who wishes to know specifically how to go about a crystal structure determination. The several methods of recording x-ray diffraction are discussed in good perspective, although there is considerable emphasis on Garrido's own method of interpreting Weissenberg photographs. The completeness of the chapter on the "Course of a structure determination" is indicated by the list of the topics covered: selection and preparation of material; determination of the symmetry; determination of the symmetry with the aid of the physical properties; determination of the symmetry with the aid of Laue diagrams; observations on the determination of the symmetry; determination of the crystallographic constants; selection of method; determination of the dimensions of the elementary cell; determination of the number of atoms contained in the elementary cell; the determination of the net type and the rules of extinction; determination of the axes of the definitive unit cell; determination of the Bravais group; elements of symmetry and extinctions; extinction tables [following Donnay and Harker]; observations on the determination of the space group; determination of the space group without utilizing x-rays [morphological determination of space group, following Donnay and Harker]; geometrical analysis of the space group; determination of the coordinates of the atoms; the Fourier method; the Fourier-Patterson method; proof of the structure.

To illustrate the determination of crystal structure by specific examples, the arguments are given for the determination of the structures of the following crystals: copper brucite, rutile, and manganite. These examples proceed

from simple to complex. The first is symmetry fixed; the second involves several possible equipoint combinations but only one parameter; the third is used to illustrate the Fourier projection of electron density, which is given in detail, even to the methods of computation; and the fourth is an example of a somewhat complex structure. Unfortunately, the chapter lacks an example of the use of Patterson or Harker methods of locating heavy atoms.

The third part of the book is concerned with a résumé of crystal chemistry as revealed by crystal structure analysis, and also by the application of x-ray methods to the study of preferred orientation in aggregates.

In the reviewer's opinion, this is one of the most comprehensive books which has appeared on the subject of crystal structure analysis. From the student's viewpoint, it is unfortunate that an English version of the work is not available.

M. J. Buerger,
Massachusetts Institute of Technology

The Chemistry of Portland Cement

By Robert Herman Bogue. Pp. 572, 6×9 in. Reinhold Publishing Corporation, New York, 1947. Price $10.00.

This excellent book is written primarily for the cement chemist, but the presentation is such that it is equally profitable reading for the cement manufacturer, the designer of concrete mixes, and for the designer and builder of plain or reinforced concrete structures. Also, it is a good reference book for students in courses on materials of construction. A distinguishing feature is that while the book is authoritative and complete, it can be read and understood by a layman with only an elementary knowledge of chemistry. The opening chapter contains an interesting history of the cement industry, and throughout the book there is a constant review of the development of the various branches of knowledge of portland cement. The viewpoint of the research chemist is apparent in the frequent references to contributions made by various investigators and in the summaries of research methods and results obtained. The references at the end of each chapter add up to a total of about 850 for the entire book.

The book is divided into three approximately equal parts with ten chapters in each part and with the following sub-titles: Part I—"The Chemistry of Clinker Formation," Part II—"The Phase Equilibria of Clinker Components," Part III—"The Chemistry of Cement Utilization."

The first part is concerned chiefly with the constitution of portland cement and with high temperature reactions in the dry state. There is a discussion of the manufacturing process together with a consideration of heat treatment in manufacture and of the design and control of cement composition. Early experimentation on the constitution of portland cement is reviewed, and more recent methods of investigation are described more or less in detail. The principal constituents of clinker are discussed, and the methods used by various investigators for detecting their presence are given.

In the second part there is a preliminary explanation of the principles and technique of high temperature phase research followed by detailed studies of the phase relations for a large number of combinations or systems of clinker components. Temperature-concentration diagrams are used freely to illustrate phase relations. This part of the book may be somewhat difficult for the lay reader, but it can be mastered by one sufficiently interested to make the effort.

The third part deals mainly with reactions between cement and water, and is the part of greatest interest from the engineers point of view. There is a review of early studies of cement hydration followed by a discussion of more modern concepts of the interaction of cement compounds with water. The chapters on heat of hydration, setting and hardening of cement pastes, reactions with sulfate solutions, and tests for cement quality should be of practical value to the designer of concrete structures.

Eugene Mirabelli
Massachusetts Institute of Technology

Physical Methods of Organic Chemistry, Vol. II

Edited by Arnold Weissberger. Pp. 1367, 16×23½ cm. Interscience Publishers, Inc., New York, 1946. Price $8.50.

The second volume of *Physical Methods of Organic Chemistry* is now available. The subjects covered in this volume and the authors are:

Spectroscopy and Spectrophotometry, by W. West
Colorimetry, Photometric Analysis, and Fluorimetry, by W. West
Polarimetry, by W. Heller
Determination of Dipole Moments, by C. P. Smyth
Conductometry, by T. Shedlovsky
Potentiometry, by L. Michaelis
Polarography, by O. H. Müller
Determination of Magnetic Susceptibility, by L. Michaelis
Determination of Radioactivity, by W. F. Bale and J. F. Bonner, Jr.
Mass Spectrometry by D. W. Stewart

A subject index for both volumes has been included in Vol. II.

This series was prepared to assemble in one place information on the physical methods of treating organic chemical problems that previously was available only in periodicals and a few specialized books. It thus aids a chemist or student who wishes to review or to understand any of the methods described. Where it was considered necessary, a discussion of the theory underlying the method is given. In some cases, such as the chapter on polarography by Otto H. Müller, this is rather complete. Other chapters are devoted to a more thorough description of the apparatus and techniques of measurement. The application, sensitivity, and accuracy of the methods, as well as the precautions which must be taken to secure precise results are given in most of the chapters. A study of one of the methods as given in this collection will permit the reader to understand and evaluate experimental results.

The practicing expert in any of the individual fields of study will probably find that some of the specific techniques that he has learned for his own work have been omitted from this series. This is to be expected in any general work of this nature. It was not the intention of either the editor or the authors to develop specialists in the various physical techniques for studying organic chemical problems. Such development is properly beyond the scope of a book of this nature. If the reader desires, after reviewing a method in this book, to make use of the techniques described, he will probably wish to utilize the supplementary literature on that particular method. For this purpose both specific references, included as footnotes, and a bibliography of general references are given in each chapter. These references constitute a valuable summary of the published work on each of the included subjects.

As is expected in a collaborated work, the presentation of the subject matter is not uniform. The disadvantage of this is outweighed by the advantage of having each topic presented by an expert in that field. In spite of the varied nature of subjects and authors, a remarkably consistent high quality of presentation is maintained throughout the series.

The editor, authors, and publishers of this book are to be congratulated on the completion of this work. It should be very useful to many workers and students in the growing field of research that involves the physical study of organic compounds.

T. J. Walsh,
Case Institute of Technology

Fundamental Theory

By A. S. Eddington. Pp. 222, 7¼×10½ in. Cambridge University Press, Teddington, England and The Macmillan Company, New York, 1947. Price $6.00.

This posthumously published book by the late Sir Arthur Stanley Eddington (1882–1944) gives us the final formulation of the theory of relativistic quantum mechanics which he made his principal study during the later years of his life. Eddington was nothing if not original, and the reader will find this book, like all of his others, filled with his unconventional points of view towards both relativity theory and quantum mechanics. It is as yet too early to assess the position which Eddington's ideas may come to play in the development of theoretical physics. It seems to have been the case, however, that his work has been so peculiarly his own that it must consequently suffer from lack of adherents who can either carry it forward themselves, or can explain it to others. It is unfortunate for the reader that the eminent editor of this volume, Sir E. T. Whittaker, did not venture his hand at bringing Eddington's treatment into clearer relation with more conventional theory, but in any event he has earned our thanks for making Eddington's final manuscript available for study.

Whether one wishes to judge Eddington's theory by its numerical predictions which can be compared with experiment, or whether he wishes to emphasize the background of the theory itself as the main thing, is probably a matter of choice. In the tradition of Eddington's own ideas, which he has so often expressed in his books, one would say that "the theory is the thing," but one senses here more than in his earlier works, that Eddington wishes his ideas to be judged heavily on their predictions. Among the more definite of these predictions, and one which is allowed to be a critical test, is the conclusion that the "nuclear force" between two protons must be of the form $A\exp(-r^2/k^2)$ with $A=4.2572\times10^{-5}$ ergs and $k=1.9208\times10^{-13}$ cm. It is unfortunate that a number of the other numerical values which come out of the theory are clouded by the fact that they involve various powers of the mysterious factor $\beta=137/136$, whose elusive characteristics are reminiscent of those of the famous snark. The applications of the theory to the problem of nuclear structure are of particular interest these days, and the reader will find them listed on page 211 of the book. These are largely binding energy studies, and to the writer it seems unfortunate that Eddington himself did not undertake to apply his methods to scattering problems; the brief discussion of radiation scattering in sections 122–3 is too closely allied to conventional arguments to present anything new.

After a lengthy tussle with this book, as well as with the earlier volume *Relativity Theory of Protons and Electrons*, which will be needed for reference in reading the present work, the writer is of the opinion that the future applicability of Eddington's work depends on whether it can be taken up by others in such a way that its development becomes possible by a number of people. If its concepts prove essentially incommunicable, then there seems to be little doubt that, despite its apparent success in Eddington's hands, it will remain simply as a monument to a mind whose very originality defied either comparison or emulation. Let us hope that the stimulus of reading this book will bring to other minds the ability to "explain" Eddington to us.

Edward L. Hill,
University of Minnesota

Applied Elasticity

By John Prescott. Pp. 666+vi. Dover Publications, New York, 1946. Price $3.95.

The present book is an unchanged reprint of the well-known 1924 edition which was published at a considerably higher price by Longmans, Green, and Company. The author's aim has been "to see the subject from the point of view of the engineer rather than from that of the mathematician. (The) theory has been developed only so far as it seemed likely to lead to the solution of practical problems."

In reviewing the present book it is well to remember that the original date of publication is 1924. The only English text at that time was Love's famous treatise. "Drang und Zwang" by A. and L. Föppl had appeared

only a little while earlier and very likely was not known to the author. Thus it was indeed the filling in of a very considerable gap which the author accomplished by offering this book. In keeping with the author's intention, the book was written in lively, informal, and direct style designed to appeal to the theoretical engineer. In the analytical portion, enough intermediate steps are included to make it relatively easy to follow along. Very many interesting examples of the general theory were worked out in full.

In addition to these positive comments it is necessary to make some negative comments. Several of the different topics in the book are treated in an *ad hoc* manner so that the connections between them and the general theory are less apparent than they might be. There is an almost complete lack of references which is emphasized by the few, not particularly significant, references which are included.

The book begins with two well written chapters on the analysis of stress and strain. Considerable space is devoted to static and dynamic problems concerning the stretching, bending and twisting of thin straight rods. As an introduction to the theory of curved rods, there are treated some problems concerning spiral springs, and the bending of circular rings in their own plane is treated fully. A subsequent chapter deals with exact solutions of the equations of elasticity for cylinders and spheres. Several chapters are concerned with the stretching and bending of flat plates. For generalized plane stress Airy's stress function is introduced, and a good many of the standard problems for circular boundaries are solved. This is followed by a treatment of the small deflection theory of transverse bending. The differential equation of the theory and the stress strain relation for plates of uniform thickness are obtained after considerable labor. Examples of application include Navier's solution for the simply supported rectangular plate and many cases of axisymmetrical bending of circular plates. The basic equations for the plate of uniform thickness are generalized, incorrectly, for plates of gradually varying thickness. (This error might have been avoided by deriving the equations of the theory by variational methods. The more so as the author subsequently gives the correct expression for the potential energy of the bent plate.)

The following chapter on *finite* small deflections of plates is interestingly written and contains some illuminating examples not to be found elsewhere.

The book continues with a chapter on the buckling of plates which includes much information on the buckling of deep beams, a fragmentary chapter on the bending and buckling of circular cylindrical shells, a chapter on transverse vibrations of rotating disks (including disks of variable thickness). The final chapter deals with Hertz' theory of elastic contact.

The reviewer feels that the manner of presentation and part of the contents of this book are sufficiently different from the more recent books of Timoshenko, Biezeno and Grammel, and Sokolnikoff, to justify the present photo-offset reprint edition. Surely, that this is so, is to the credit of the author's pioneering effort which must have brought the theory of elasticity for the first time within the reach of many English-speaking engineers.

E. Reissner,
Massachusetts Institute of Technology

The Rare-Earth Elements and Their Compounds

By Don M. Yost, Horace Russell, Jr., and Clifford S. Garner. Pp. 92+viii. John Wiley and Son, Inc., New York and Chapman and Hall, Ltd., London. Price $2.50.

This book gives an excellent review of the chemical and physical researches which have been carried out using rare earths. It is well organized, authoritative, and an extremely useful book for anyone wishing to do further work involving rare earths. The chapters on electronic structure, oxidation states of the rare earth elements, paramagnetic properties of the rare earth compounds, the absorption spectra of rare earth compounds, and the chemical and physical properties of the rare earths, are excellent.

The chapter on the separation of the rare earths is good but unfortunately was written during the war period, so that it does not contain an account of the extensive research work in this field carried out during the war. Particularly, the excellent work of Marsh and his co-workers on the amalgam separations of the divalent rare earths (J. Chem. Soc. 398–401 (1942); *ibid.* 523–26 (1942); *ibid.* 8–10 (1943); *ibid.* 531–5 (1943)) are not emphasized. Also, the extensive investigations on ion exchange separations of the rare earths which were just recently released in a symposium at the meeting of the American Chemical Society in New York were not available at the time the book was written. Similar lack is to be noted in the listing of the nuclear properties, since much of this information, although very extensive, is just now becoming available.

This book would be a very valuable one in anyone's physical or chemical library because of the widespread interest in the rare earths which has been aroused by the fact that they occur among the fission products of thorium and uranium and by the fact that many of these elements are now available because of the rapid separation methods which can be carried out by means of the ion exchange resins.

F. H. Spedding,
Iowa State College

Currents in Aerials and High-Frequency Networks

By F. B. Pidduck, Pp. 97, Illustrations 29. Oxford University Press.

Several methods of antenna analysis are based on integral equations of the first kind. As early as 1897 Pocklington derived an approximate equation of this type for a thin wire and used it to show that in the first approximation the current in the wire is distributed sinusoidally. The conclusion is now well known, although the original paper

attracted but little attention. In 1931 Murray obtained exact equations of the same type as Pocklington's and proposed a method of solution by means of orthogonal expansions, such as Fourier series. This method and its applications are developed in detail in *Currents in Aerials and High-Frequency Networks* by F. B. Pidduck. Known and unknown functions are expressed as Fourier series, and an infinite system of linear algebraic equations is obtained for the unknown coefficients. In principle the method is very simple and general; in practice it is distinguished by a forbidding complexity. Thus, at the very beginning the author makes approximations in his integral equations and restricts himself for the most part to nearly resonant antennas. While the equations are still complex, he succeeds in obtaining several concrete results.

The treatment of the subject could have been more rigorous. In those methods of antenna anlysis which involve integral equations, mathematical rigor is not merely an academic matter. Seemingly reasonable approximations in integral equations may render them solutionless. Following Pocklington, the author replaces the integrals arising from the surface current distributions by line integrals taken along the axes of the wires. The physical situation described by the altered equations happens to be an impossible one so that one would not expect these equations to possess solutions; and it can be shown mathematically that this is indeed the case. Engineers habitually make similar approximations but under circumstances which as a rule do not involve logical contradictions. In the case of a very long wire the author arrives at the startling conclusion that the current is constant along the wire except near the ends, where it sinks to zero "in a distance comparable with its diameter." This conclusion is a consequence of improper approximations in the distant terms of an infinite series. A similar error is made in the analysis of diffraction by a conducting half-plane.

No logical difficulties would be encountered in applying the method to Murray's original integral equations, but convergence questions would have to be treated seriously. The difference between Pocklington's approximate equations and Murray's exact equations manifests itself significantly only in the high order terms of Fourier series. This is a warning of troubles to come. Similarly, when the impressed electromotive force in a transmitting antenna or the load in the receiving antenna is highly concentrated, the length of the interval for this distribution may be very important under some conditions but this length does not enter the equations, except in the high order coefficients of Fourier series. Thus, when the basic difficulties of convergence are overcome, those of summation will have to be faced before reliable concrete results can be obtained.

I do not share the author's enthusiasm for Murray's method, but perhaps it is too early to pass the final verdict. In this book the author has done much to reduce this method to practice; but it still remains for someone to adapt the method to exact integral equations, to discover a convenient method of summation, and to extend the application to non-resonant antennas.

S. A. SCHELKUNOFF,
Bell Telephone Laboratories, Inc.

New Booklets

National Bureau of Standards, Washington 25, D. C., has announced the publication of a complete listing of commercial and university testing and research laboratories throughout the country, together with indications of the type of commodities tested. This pamphlet, designated as NBS Miscellaneous Publication M187 and entitled *Directory of Commercial and College Laboratories*, is available only from the Superintendent of Documents, Washington 25, D. C., at 30 cents per copy.

The RCA Tube Department of Radio Corporation of America, Camden, New Jersey, has announced the publication of a new edition of the RCA Receiving Tube Manual, RC-15. This edition is the first since 1939. 35 cents. Available from RCA tube distributors, as well as from the home office in Camden, New Jersey.

Tracerlab, Inc., 55 Oliver Street, Boston, Massachusetts, announces issue No. 7 of its monthly publication, *Tracerlog*, which features radioactivity measurements. 12 pages. Free on request.

Allied Radio Corporation, 833 West Jackson Boulevard, Chicago 7, Illinois, has published a new 48-page supplement to its regular master catalogue. The supplement, No. 114, as well as the regular catalogue, No. 112, can be obtained without charge.

Harry W. Dietert Company, 9330 Roselawn Avenue, Detroit 4, Michigan, recently issued two new booklets which are available on request: a four-page leaflet entitled "Moistmeter," describing a new portable electric moisture determinator, and a 12-page brochure on carbon and sulfur analysis.

Executone, Inc., 415 Lexington Avenue, New York 17, New York, has published a booklet, "Sound—A Modern Control System," which presents the company's complete line of voice-paging and music systems. 8 pages.

Here and There

New Appointments

Thomas G. Digges has been appointed chief of the Thermal Metallurgy Section of the National Bureau of Standards, succeeding D. J. McAdam, who retired August 31, 1947, after being chief since 1930.

Harold H. Loudin, formerly a research physicist with the Owens-Corning Fiberglas Corporation, is now an Assistant Professor of Physics at Kent State University.

Otto Halpern was recently appointed Professor of Theoretical Physics at the University of Southern California.

Transfer of Geophysical Abstracts

The United States Geological Survey has resumed publication of the Geophysical Abstracts after a four-year interval, during which they were issued by the U. S. Bureau of Mines. The Abstracts are published quarterly as an aid to those engaged in geophysical research and exploration. Copies may be purchased singly or by annual subscription from the Superintendent of Documents, Government Printing Office, Washington 25, D. C. For subscription, the Superintendent of Documents will accept a deposit of $5 in payment for subsequent issues. When this fund is near depletion the subscriber will be notified.

American Instrument Fair

The Third Instrument Conference and Exhibit will be held in Convention Hall in Philadelphia, September 13–17, 1948. The Exhibit will be known as the "American Instrument Fair."

New Hebrew Publication

A new journal entitled *Scientific Publications* is being published in Haifa, Palestine, at Hebrew Technical College. Volume 2 contains the following papers printed in Hebrew, with an English summary included:

An elementary derivation of the equivalence of mass and energy, by Albert Einstein;
The spreading of mercury in tin foils, by A. Tcherniavsky;
Vision as a technical problem, by F. Ollendorff.

National Research Council Fellowships

The National Research Council has announced that it is again offering a number of opportunities for advanced study and research in the natural sciences. Fellowship boards established by the Council will administer fellowship programs financed by Merck and Company and Radio Corporation of America. The postdoctoral fellowship program administered by the Council for the past twenty-eight years will continue. Further particulars concerning these programs may be obtained upon request from the NRC office at 2101 Constitution Avenue N. W., Washington 25, D. C.

Award of Optical Medal

William F. Meggers, chief of the spectroscopy section of the United States Bureau of Standards, was awarded on October 24 the Frederick Ives Medal of the Optical Society of America for distinguished work in optics.

The Bulletin of Mathematical Biophysics

Following is the table of contents of the December 1947 issue of *The Bulletin of Mathematical Biophysics* (The University of Chicago Press, Chicago, Illinois. Volume 9, Number 4.):

An approach to a neurobiophysical interpretation of motivational interactions—H. D. Landahl
Contributions to the theory of organic form: the intestinal tract—Alfonso Shimbel
Suggested experimental procedure for determining the satisfaction function of animals—Anatol Rapoport and Alfonso Shimbel
Theory of ameboid movements—N. Rashevsky
Book review—Alston S. Householder—*Mathematical Analysis of Binocular Vision* by Rudolf K. Luneburg

Scientific Books for Finland

Arthur E. Morgan, member of the American Friends Service Committee, Yellow Springs, Ohio, has made a plea for contributions of scientific and technical books and periodicals from America to take the place of those destroyed in Finland in the recent war. The library of Finland's Technical Institute was bombed and totally destroyed. Dr. Morgan writes: "In the remarkable efforts for recovery that the Finns are making, the lack of technical library facilities is a very serious handicap. It would be a practical act of friendship to a nation that holds America in high regard if Americans should contribute good technical books and periodicals to this library." Gifts should be marked for the Institute of Technology, Helsinki, and sent to the Legation of Finland, 2144 Wyoming Avenue N. W., Washington, D. C. Dr. K. T. Jutila, the Finnish Minister, will arrange for their being shipped to Finland.

1948 Conference of Lighting Engineers

The next annual meeting of the Illuminating Engineering Society will be held at the Hotel Statler in Boston, September 20–24, 1948.

New Statistical Journal

In August 1947 was published the first issue of *The American Statistician*, which replaces the *ASA Bulletin* as the news publication of the American Statistical Association. Non-members may subscribe at the annual rate of $1.50; single copies cost 25 cents. The address of the Association is 1603 K Street N. W., Washington 6, D. C. The table of contents of the first issue follows:

A welcome to the international statistical conferences—W. Averell Harriman,
A. S. A. programs arranged for the international statistical conferences,
Statistical engineering—Robert C. Tumbleson,
Lemuel Shattuck, Statist—Walter F. Willcox,
New high-speed computing devices—Franz L. Alt,
A well-rounded curriculum in statistics—William A. Neiswanger and H. K. Allen, and
Hold that line—a plea for the preservation of chart scale rulings—Kenneth W. Haemer

Metallurgy Meeting of Interest to Physicists

The American Institute of Mining and Metallurgical Engineers has announced the special series of joint sessions on theoretical physical metallurgy which will be included in the annual meeting at the Pennsylvania Hotel, New York City, February 16th–19th. Papers which should be of interest to solid-state physicists will be delivered in the morning and afternoon sessions on February 18th and 19th. For information as to program, write the above Institute at 29 West 39th Street, New York 18, New York.

Proceedings of the Electron Microscope Society of America

THE annual meeting of the Electron Microscope Society of America was held at Franklin Institute Auditorium in Philadelphia, Pennsylvania, December 11–13, 1947. Titles and abstracts of the papers presented are given below:

1. Electron Microscopy of Keratin Fibers. J. L. FARRANT,* E. H. MERCER,† AND A. L. G. REES,* *Council for Scientific and Industrial Research* (to be read by J. M. COWLEY, C.S.I.R.).

Electron microscopical studies of the Keratin fibers, the most important of the natural fibrous proteins, have contributed much to their general histology. Our earlier work demonstrated the physical and chemical differentiation of various constituent parts of the fiber and this has been substantiated recently by the direct chemical investigations of Lindley. Application of techniques such as metal shadow casting indicates that the intracellular matrix of the cortical cells is an amorphous protein having molecules ~110 A.U. in diameter and that the fibrous part of the cortex is composed of protofibrils about 110 A.U. wide and ~0.3μ in length twisted together in a rope-like manner to form fibrils of diameter ~0.2 to 0.4μ extending along the cortical cells. These protofibrils are further shown to be linear aggregates of approximately isometric particles ~110 A.U. in diameter. This evidence, which has been exhaustively tested to eliminate the possibility of it being of artifact origin, is in contradiction to Astbury's conception of the molecular chains extended parallel to the fiber axis in such proteins and indicates a closer association with the globular proteins. The implication of the electron microscopical evidence, together with that obtained from x-ray and chemical studies on these proteins, are discussed.

* Div. of Industrial Chem., Melbourne, Australia.
† Div. of Physics, Sydney, Australia.

2. An Investigation of Degraded Wool. GLORIA SEEMAN AND MAX SWERDLOW, *National Bureau of Standards, Washington, D. C.*

The effects of various physical and chemical factors leading to deterioration of fabrics were investigated by the electron microscope. Wool fibers from treated and untreated fabrics were studied by a thermoplastic replica technique so that the fibers are not changed during preparation and are available for further treatment. A study of the effects of mechanical abrasion on woolen fabrics has also led to some observations of the morphology of fibers that can be correlated with the observations of other investigators. The techniques employed will be described.

3. Electron Microscope Studies of Nerve Axon Structure. E. DE ROBERTIS AND F. O. SCHMITT, *Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

Unmedullated fibers from the lobster, horseshoe crab, squid, and interganglionic segments of human sympathetic chains and medullated fibers from frog and rabbit sciatics and bull frog motor roots were examined after fragmentation of fresh and formalin-fixed nerves. In all these types a new fibrous structure was observed, having indefinite length and widths ranging from 200 to 1100A, depending on nerve type. The chief features are the high density of the edges, low density of the central region and a periodic cross banding. The periodic axial structure is seen best after washing and staining. In sympathetic nerves the axial period is about 650A. Three intraperiod bands, having characteristic density and position, have been observed. The structural and chemical properties differ distinctly from those of collagen, which has the same axial period. An amorphous dense material, removable by water washing, is associated with the structures. These fibrous structures were obtained from isolated, fixed axons of squid giant fibers, hence they occur in the axis cylinder. All available evidence is consistent with the view that the structures are tubular in nature. If this is confirmed by adequate examination of the shape of their cross sections they will be referred to as "neurotubules."

4. Polymorphism of Organic Pigments. F. A. HAMM, *General Aniline and Film Corporation, Easton, Pennsylvania.*

Organic pigments are often recrystallized from sulfuric acid to effect purification and to control the particle size of the new crop of crystals. Copper phthalocyanine (Heliogen Blue) and Indanthrene Blue RS have been observed to yield crystals of varying shades, ranging from a light greenish blue to a dark reddish blue (purple). These shades have in the past been attributed only to differences in particle size because the chemical activity and elementary analyses strongly indicated that the compounds were chemically identical. This investigation shows that both these pigments are polymorphic, and each can exist in at least three different physical structures. The shade differences are, therefore, very probably due to the differences in the light absorption exhibited by the structurally different polymorphic forms. The polymorphic transformations can, for both pigments, be readily observed to take place in the electron microscope. The new phase, stable during and after exposure to the illuminating electron beam, can be seen to grow from the vapor state at the expense of the original solid material. Electron diffraction patterns prove that the newly formed phases are structurally different from both the original samples. However, both "red" and "blue" shades of each blue pigment transform to give the same new structure. Both these commercially important pigments are therefore at least trimorphic.

5. Structure of Cellulose Fibers as Revealed by the Electron Microscope. CHARLES W. HOCK, *Hercules Experiment Station, Hercules Powder Company, Wilmington, Delaware.*

The structure of cellulose fibers was studied advantageously by means of metal-shadowed specimens, surface

replicas, and stained preparations. These techniques of preparating specimens for electron microscopical examination clearly revealed the fundamentally fibrillate structure of natural cellulose, as well as other details not readily discernible in specimens prepared by ordinary techniques.

6. Replica Studies of Dyed Nylon. F. A. HAMM, *General Aniline and Film Corporation, Easton, Pennsylvania.*

The crystal size of pigments and dyes when applied to textile fibers is almost invariably below the limit of light microscopical resolution. The dyestuff chemist has, however, had an insatiable desire to learn something about the distribution of his dyestuff on these fibers. This paper describes a technique for preparing replicas of nylon dyeings containing crystals ranging from about 100 to 5000 A.U. long. These replicas are prepared by subliming the silica directly onto the nylon fibers, followed by a mild acid extraction of the nylon. Two types of replicas are possible: (1) those consisting of silica only, and (2), those consisting of silica with the dye crystals themselves in the same relative position as they were on the nylon fiber. The second type is obviously more informative because of the greater contrast (resolution) and because fewer artifacts are possible. This technique has been successfully used on nylon dyed with such commercially important vat dyes as Indanthrene Golden Yellow IGK, Indanthrene Brown IRRA, and Indanthrene Brilliant Pink IRA. The tremendous crystal size range, their shape, and their orientation on the nylon surface are of interest. Striations on the surface of the nylon fibers are clearly visible. Similar investigations on other fibers are being carried out.

7. The Keysort Bibliography of Electron Microscopy. W. B. HOWE, *The McBee Company, Collingswood, New Jersey.*

The new bibliography of electron microscopy on Keysort cards is described and the Keysort method of filing and sorting is explained. Methods of punching author, subject, and source indices are suggested. A tentative subject index which will permit workers in various fields to expand the sorting system in their special fields is described.

Symposium—Teaching of Electron Microscopy

8. Introductory Remarks. R. C. WILLIAMS, *University of Michigan, Ann Arbor, Michigan.*

9. University Courses in Electron Optics and Electron Microscopy. S. G. ELLIS, *RCA Laboratories, Princeton, New Jersey.*

10. The Teaching of Electron Microscopy and Its Role as a Research Method in Biology. C. E. HALL, *Massachusetts Institute of Technology, Cambridge, Massachusetts.*

11. The Training of Technical Electron Microscopists. P. C. SMITH, *RCA Victor Division, Camden, New Jersey.*

12. What Does Industry Expect of a Trained Electron Microscopist? J. H. L. WATSON, *Henry Ford Hospital, Detroit, Michigan.*

13. Qualifications of an Industrial Electron Microscopist. W. G. KINSINGER, *Hercules Powder Company, Wilmington, Delaware.*

14. Recent Developments in the Shadow-Casting Process. ROBLEY C. WILLIAMS AND ROBERT C. BACKUS, *Department of Physics, University of Michigan.*

The electron micrography of extremely small organic objects is rendered uncertain by the surface structure of the plastic-film substrates commonly used. Polished glass has a remarkably smooth surface, and excellent pictures have been obtained by the transfer-replica process, with gold used as the shadow-casting metal. However, a high intensity, biased-beam, electron gun frequently causes a serious granulation of the gold film. Attempts to remedy this defect have led to a re-examination of several other substances as shadowing agents. Most metals which do not granulate under electron bombardment adhere so tightly to a glass surface that they are unsuitable for the transfer-replica method. The metals which readily oxidize, such as chromium and uranium, are found to adhere tightly to glass, while non-oxidizing ones are readily stripped. Uranium is the ideal shadowing material, however, and its sulfide is found to be non-adherent to glass, and non-granulating under electron impact. Uranium sulfide may be used directly, or it may be created from a film previously deposited as uranium. The process which appears best to permit micrography of minute objects is the transfer-replica one, with glass as a substrate, and uranium sulfide as the shadowing material.

15. Modified Silica Replica Technique for Study of Biological Membranes and Rotary Shadow Casting of Specimen. F. HEINMETS, with the technical assistance of J. T. QUYNN, *Chemical Corps, Camp Detrick, Frederick, Maryland.*

A discussion is given on the use of the silica replica technique for the study of the structure of biological membranes. Electron micrographs of dried chick chorioallantoic membranes are presented, together with other illustrative evidence. Rotary shadow casting of replicas and other specimens will be discussed.

16. The Range of Migration of Silica Molecules in the Formation of Silica Replicas. C. J. CALBICK, *Bell Telephone Laboratories, Inc.*

Silica condensing upon a surface at an oblique angle of incidence produces shadows of obstacles such as particles

present on the surface, similar to those produced by the metal shadow-casting technique of Williams and Wyckoff. In contrast to the behavior of heavy metals, silica diffuses from the regions of direct incidence into the shadows, there forming a film thin relative to that in regions directly exposed to condensing silica. Densitometer traces of electron micrographs show: (1) uniform thickness far from any shadow, (2) slowly decreasing thickness as a shadow edge is approached, (3) a sharp decrease in thickness across the shadow edge, (4) slowly decreasing thickness within the shadow as distance from the edge increases. This behavior can be explained by assuming that high temperature silica molecules condensing upon a surface at room temperature have a probability for inelastic collision less than unity, and the reflected silica is trapped in a surface layer to form a "two-dimensional gas" which diffuses over the surface. Whenever the molecules composing this gas collide with the surface they have a definite probability of sticking; consequently, the film within the shadow grows thinner as one recedes from the shadow edge. The densitometer measurements yield values for the probability for inelastic collision (0.54), the range of migration, i.e., the distance within which half of the silica diffusing from a shadow edge condenses on the surface (0.53μ), and the average distance a molecule travels between collisions with the surface (0.42μ). The bearing of these results upon the mechanism of formation of silica replicas will be discussed.

17. A Positive-Replica Technique for Electron Microscopy. C. M. SCHWARTZ, A. E. AUSTIN, AND P. WEBER, *Battelle Memorial Institute, Columbus, Ohio.*

In the course of an investigation, now in progress,* of the effect upon the surface contour of bearings run under conditions of fluid film lubrication, the well-known shadow-cast plastic film negative replica appeared inadequate, owing to the fact that the ridges in the replica (which are the regions highlighted and enhanced by shadow casting) correspond to valleys in the original surface, rather than to the high spots. The latter are of utmost importance to this study, if metal-to-metal contact occurs. The need arose, therefore, for a positive-replica technique, suitable for shadow casting. Polyvinyl alcohol (PVA) has been found quite suitable as an intermediate negative replica, and, in combination with formvar, each satisfies the condition for mutual insolubility in the solvent for the other. The PVA replica has been prepared by impression of sheet material, using moderate heat and pressure, and by application of a water latex which is dried under a heat lamp. A formvar solution in ethylene dichloride has been used to produce the positive replica; this is readily removed from the PVA by floating the composite film on water, PVA side down, whereupon the latter dissolves. Washing is completely effected by continuous dilution, and the formvar film subsequently handled in the usual manner. The structural details do not appear to have suffered through the replica reversal. The technique has been applied to a study of the surface changes occurring during bearing wear.

* ONR Contract No. N5-ORI-111, Task Order III.

18. An Interferometric Method for the Calibration of Electron Microscope Magnification. J. L. FARRANT AND A. J. HODGE, *Division of Industrial Chemistry, Council for Scientific and Industrial Research, Melbourne, Australia* (to be read by J. M. COWLEY, C.S.I.R.).

Interferometric methods have been used to measure uniform cylindrical silica and glass fibers as small as 5μ in diameter with better than 1 percent accuracy. Such fibers yield sharp images when examined in the electron microscope and constitute a convenient and rugged specimen for accurate direct calibration of the instrument for magnifications up to 10,000 diameters. Higher magnification ranges are then readily determined by means of an internal calibration obtained by taking micrographs of glass spheres about 0.5 micron in diameter. Such spheres are also convenient specimens for determining the image distortions which must be known to correct the widths of the fiber images. This method appears to be capable of greater accuracy than any employed hitherto.

19. The Hydrogen-Diffusion Leak Locator. HERBERT NELSON, *RCA Victor Division, Harrison, New Jersey.*

A new instrument has been developed for locating leaks in vacuum devices and vacuum equipment. It employs a sealed-off, highly evacuated ionization gauge having a section of its envelope made of thin palladium sheet which, when heated, is highly permeable to hydrogen. In the instrument, the gauge is connected to an all-metal manifold with the palladium section isolating the vacuum of the gauge from that of the manifold. When a leaky vacuum device connected to the manifold is probed with hydrogen, leaks are indicated by an increase in the ion current of the gauge. Since gases or vapors other than hydrogen do not penetrate the palladium, high vacuum is not required in the manifold nor in the device under test. A vacuum of the order of 1 to 10 microns is sufficient. The new instrument is similar to the mass-spectrometer leak locator in that it owes its high sensitivity to its ability to sort out, from other gases and vapors, the tracer gas (H_2) which enters through the leak. Thus, the palladium window in the hydrogen-diffusion leak locator serves the same purpose as does the crossed electric and magnetic fields of the mass-spectrometer leak locator.

20. The Electron Microscopy Group of the British Institute of Physics: Its Constitution and Activities. V. E. COSSLETT (Hon. Secretary) (to be read by A. S. MCFARLANE).

The origin of the Group was the informal meeting of users of the RCA electron microscopes sent to Britain under the lend-lease scheme. Some account will be given of its earlier conferences and its present constitution as a constituent part of the Institute of Physics, with 80 inscribed members. Current activities include compilation of a bibliography of the subject, preparation of abstracts, formation of a collection of reprints for the use of members, and plans for a Vacation School in electron microscopy. Meetings for the reading of papers are held twice yearly. A list will be presented of the 32 laboratories in Britain

now operating electron microscopes, and of their respective fields of interest.

21. European Electron Microscopes. R. W. G. Wyckoff, *National Institute of Health, Bethesda, Maryland.*

During the course of a stay in Europe this last year, the author had the opportunity to see examples of all but one of the types of electron microscopes now available there and to work with several of them. This is a brief discussion of their salient characteristics and an estimate of what they can do compared with the performance of the instruments with which we work.

22. Dark-Field Electron Microscopy. Cecil E. Hall, *Massachusetts Institute of Technology.* (Invited Paper.)

Electron microscope images in dark-field were studied by means of an aperture system fixed to the object rather than the objective lens. The resolution for images produced by the diffusely scattered component was found to be in the range 100 to 200A, but the resolution in images produced by coherent scattering approaches that obtainable in bright-field. A resolution of about 50A was obtained with test objects consisting of evaporated films containing small crystallites, but it is estimated that crystallites having dimensions down to about 20A can be recorded under the conditions prevailing in this work. Numerous substances were studied in dark-field and as a result of this survey it is concluded that the dark-field method is of value for studying the size, location, and other characteristics of crystalline components in a specimen. Some of the theoretical factors influencing resolution in dark-field are discussed. As an example of the application of the dark-field method the studies of colloidal carbon may be cited. Anomalous intensities in the dark-field images of certain commercial carbon blacks are interpreted as resulting from coherent scattering by parallel layer groups of graphitic planes which were previously identified by Biscoe and Warren using x-ray diffraction. The dark-field micrographs indicate that in a non-reinforcing and a semi-reinforcing black, the parallel layer groups near the surface of the roughly spherical particles are oriented with their representative vector approximately normal to the surface. No definite evidence was obtainable for the arrangement in a reinforcing channel black with average particle diameter about 300A. After prolonged heat treatment to the point where graphitization occurs, the channel black exhibits numerous diffraction images in dark-field indicating the presence of crystals having dimensions in the range 20 to 90A.

23. The Effect of Chromatic Error on the Resolving Power of Dark Field Electron Micrographs. E. G. Ramberg and J. Hillier, *RCA Laboratories, Princeton, New Jersey.*

The electron velocity distribution introduced by a thin collodion membrane has been measured by means of the electron micro-analyzer. This permits an estimate to be made of the effect on the resolution attainable in dark field electron micrographs.

24. Further Studies on the Use of an Objective Aperture. J. Hillier, *RCA Laboratories, Princeton, New Jersey.*

The effect on an electron microscope image of the use of an objective aperture, of smaller than optimum size, has been studied. Physical apertures providing angular apertures in the range 10^{-2} to 5×10^{-4} radian were used in the experimental double objective lens system described earlier. Angular apertures near the optimum value of 5×10^{-3} radian were found to increase the gross contrast but not the detailed contrast, particularly in extremely thin specimens. Angular apertures an order of magnitude smaller, however, were found to increase the detailed contrast in all kinds of specimens. With all apertures the loss of resolving power was greater than would be expected from the Abbe formula. Some of the difficulties encountered in this regard and the attempts to overcome them are described.

25. Central Research Laboratories High-Speed Micro-Oscillograph. Gordon M. Lee, *Central Research Laboratories, Inc., Red Wing, Minnesota.*

An improved version of the high-speed micro-oscillograph previously described in the literature[1] has been developed. In addition to over-all improvement of mechanical design and workmanship, specific mechanical improvements consist of a stainless-steel vacuum system, a simple trouble-free vacuum-joint design, and integral mounting of the vacuum pumps. Deflection characteristics have been much improved resulting in negligible trapezoidal distortion, doubled sensitivity and much less defocusing of the writing spot. The resulting apparatus is an easily operated research instrument for making single-sweep records of three simultaneous phenomena at frequencies up to 10,000 megacycles.

[1] Gordon M. Lee, Proc. I.R.E. **34**, 121W (1946).

26. On the Visual Focusing of an Electron Microscope. J. Hillier and S. G. Ellis, *RCA Laboratories, Princeton, New Jersey.*

The appearance of electron microscope images at and near focus are described in detail. The various methods of visual focusing are described and their relative merits and limitations are discussed. Particular reference is made to the more recently developed methods in which a reduced final image is examined at relatively high optical magnification and in which the angular aperture of the illuminating beam is effectively increased by a deflection system.

27. A Discussion of Methods and Techniques for Particle Size-Determinations with Electron Microscopes. John H. L. Watson, *Research Institute of the Henry Ford Hospital, Detroit 2, Michigan.*

The important errors and difficulties which are inherent in electron optical procedures are to be described as they apply to particle work, and means are to be given for minimizing them. Laboratory techniques for obtaining data are to be described and convenient statistical methods explained. The uses of logarithmic and arithmetic probability graph paper will be discussed. The methods apply

to measurements made upon all types of structure and the word "particle" is used in the unrestricted sense of "the thing measured for statistical treatment." Routine employment of the electron microscope for securing semi-quantitative statistical data is to be described under the following headings:

Introduction.
Mounting Procedures.
Conditions Which Affect Measurements in Electron Micrographs.

1. Resolution.
2. Contrast.
3. Distortion.
4. Focus.
5. Magnification.
6. Specimen Changes Under Bombardment in a Vacuum.
7. Use of a Biased Electron Gun.

Procedures in Particle Counting.
Statistical Methods.
Two Special Techniques.
Examples.

1. Shawinigan Black.
2. Polyvinyl Acetate Emulsions.
3. Sperm.
4. Bacillus Friedlander.

28. The Rod Artifact in the Electron Microscope Image. L. H. WILLISFORD, *Goodyear Tire and Rubber Company, Akron, Ohio.*

The rod artifact between curved profiles is present in electron micrographs of 20A° resolution. Sharp edged images with rod artifacts 1000A° long have been obtained with the electron microscope by using poor focus and the extreme contrast afforded by small objective apertures and photographic procedures. With specimens having curved profiles near together, in contact, or slightly overlapping, the rod artifact can be used as a measure of the limit of resolution. The diffraction effects operating to produce the rod between curved profiles are, of course, operating when profiles of any contour are closer than the limit of resolution, and consideration of this fact is helpful in interpreting the true specimen structure from electron micrographs.

29. Recent Advances in Electron Microscopy in Great Britain. THE BRITISH ELECTRON MICROSCOPY GROUP (to be read by Dr. C. F. Robinow). (Invited paper.)

The chief contributions of the various British laboratories operating electron microscopes will be described: Investigation, layer by layer, of the structure of earthworm cuticle and of muscle fibers (Leeds). Bacterial morphology, life-cycle of tubercle bacilli, plant virus size and shape (including crystal structure of turnip virus), use of beryllium films, origin of contamination of specimens, structure of slip bands in plastic deformation of metals (by oxide replicas) (Cambridge). Surface studies of paint films, by a new replica method (Paint Research Station, London). Construction of a thermal analyzer of dust clouds, and comparative light and electron microscopical studies with it (Chemical Defence Station). Technique of preparation of bacterial specimens, and studies of gold aggregation under influence of the beam in shadowed specimens (National Institute for Medical Research, London). Replicas of wet specimens by a polymerization process (Oxford and Harwell). Studies of clays (English Clays, Ltd.). Profile technique for bacteria (Middlesex Hospital, London). Comparison of electron microscopical replica techniques with optical and phase microscopical examination of metallographic specimens (National Physical Laboratory). New three-stage electron microscope for the research laboratory (Metropolitan-Vickers, Ltd.).

30. Interrelations of the Electron Diffraction Techniques and Electron Microscopy. EARL A. GULBRANSEN, *Chemical Department, Westinghouse Electric Corporation, East Pittsburgh, Pennsylvania.* (Invited paper.)

31. Design Features of a New Electron Diffraction Unit. PERRY C. SMITH, *Scientific Instruments Engineering, RCA Victor Division, Radio Corporation of America, Camden, New Jersey.*

The design and salient features of a new electron diffraction camera are described. Mechanical and electrical components, as well as a new vacuum control system, are shown.

32. Operational Features of a New Electron Diffraction Unit. ROBERT G. PICARD, *Scientific Instruments Engineering, RCA Victor Division, Radio Corporation of America, Camden, New Jersey.*

The results obtained by using various combinations of lenses in electron diffraction studies are discussed. Illustrations are given of the effect of varying electron probe size and position which demonstrate the use of the instrument as a point microscope or diffraction camera. Electron optical magnification of patterns permits the unit to function as a diffraction camera with effective specimen-to-plate distances from 50 to 200 cm. Patterns from insulating materials are shown with and without the use of an auxiliary electron gun for charge neutralization.

33. The Fine Structure in the Electron Diffraction Pattern from MgO Smoke. T. MCLAUCHLAN AND S. G. ELLIS,* *Department of Physics, University of Toronto.*

An explanation of the fine structure in the electron diffraction pattern from MgO smoke is advanced on the basis of the kinematic theory of electron diffraction. The explanation differs from that of Cowley and Rees in that only a single valued inner potential is assumed. Some considerations involved in the focusing and testing of electron diffraction camera are discussed.

* Now at RCA Laboratories, Princeton, New Jersey.

34. Notes on Electron Diffraction with the Universal Electron Microscope.* R. E. DUNAWAY AND H. M. BARTON, *Research Department, Phillips Petroleum Company, Bartlesville, Oklahoma.*

By placing a specimen in the electron microscope stage and mounting a small aperture (0.015 inch) on the specimen holder between the electron source and the specimen in

the RCA Type EMU electron microscope, enlarged electron diffraction patterns may be obtained when the projector and objective lens pole pieces are removed. At proper settings of the current in these lenses, plate-to-specimen distances equivalent to as much as one meter may be obtained. These magnified diffraction patterns give no evidence of distortion, although patterns made with the specimen in the normal position for electron diffraction —below the projector lens—result in elliptical rings, on this particular instrument.

* Phillips Petroleum Company, Research Department, Report No. 647-47R.

35. Precipitation Phenomena as Revealed by the Electron Microscope. ROBERT B. FISCHER, *Department of Chemistry, University of Illinois, Urbana, Illinois.*

The study of the formation of precipitates from solution is one field in which the high useful magnification of the electron microscope is of value. This paper discusses a number of precipitation phenomena encountered in analytical chemistry. Although cobalt ion does not prevent the precipitation of nickel dimethylglyoxime, it does have an appreciable effect upon the crystal shape of the precipitate. The precipitation of bismuth dimethylglyoxime is shown to be in reality a post-precipitation of that compound upon a bismuth oxy-compound. The peculiar appearance of palladium dimethylglyoxime is somewhat different under the electron microscope than under the optical microscope. The effects of *p*H, chloride ion, nitrate ion, acetate ion, and acetic acid upon the precipitation of benzidine sulfate are all shown to be a prevention of proper crystal growth and not a mere increase in solubility of the benzidine sulfate. The coprecipitation of ferric ion and of nitrate ion with barium sulfate is generally considered to be an absorption on the crystal surfaces of primary and counter ion layers and an occlusion of these ion layers as small crystals aggregate; that this is not the full mechanism is shown by electron micrographs of protuberances many ions thick on otherwise smooth barium sulfate crystals.

36. A Study of Gold Smokes with the Electron Microscope.* LOUIS HARRIS, *Massachusetts Institute of Technology, Cambridge, Massachusetts,* AND BENJAMIN M. SIEGEL,** *Polytechnic Institute of Brooklyn, Brooklyn, New York.*

The particle size and the distribution in deposits of various gold smokes have been examined at high resolution with the electron microscope. The smokes were produced by evaporating gold from a hot tungsten filament in a partial atmosphere of nitrogen and were deposited on thin collodion films. It was found that both the size of the unit colloidal gold particles and the manner of their aggregation varied markedly with the pressure of the atmosphere and the rate of evaporation. The optical properties of the deposits were also found to depend upon these factors. Measurements in the visible and infra-red regions of the spectrum have been correlated with the electron microscope observations.

* This work has been supported by the Navy Department, Bureau of Ships, Contract NObsr 39069.

** Associate of Weizmann Research Center, Rehovoth, Palestine.

37. Selected Applications of the Electron Microscope in Ceramic Research. ROBERT B. FISCHER, *Department of Chemistry, University of Illinois, Urbana, Illinois.*

The general utility of the electron microscope in ceramic research is briefly surveyed. A few recent specific applications are discussed, including the investigation of the heat-induced transformation of kaolinite to mullite, and a correlation of the size and shape of enamel pigment particles with gross physical properties of the enamel.

38. A New High Speed Microtome. HARVEY B. MERRILL, *Custom Scientific Instruments, Inc., Arlington, New Jersey,* AND ERNEST F. FULLAM, *General Electric Company, Schenectady, New York.*

The high speed microtome having proven its value in cutting thin cross sections for both the light and electron microscopes, more attention has been given to the improvement of its design and its ease of operation. The knife disk spindle has been mounted independent of the motor for ease of adjustment. A new knife disk has been designed which clamps a specially made, long-ridged knife to provide a longer working edge, higher maximum edge velocity, and most important, freedom from knife vibration. The blade is quickly interchangeable on the disk, yet is far more securely attached. Other improvements have been made, particularly in the sample feed and section collecting mechanisms and disk shield. The control panel has been changed to provide better regulation and protection of the cross feed and disk motors by means of a time delay relay operating a vacuum tube rectifier circuit and a timer switch operating through the variable transformer of the disk motor. The wiring has also been changed to reduce the complexity and size of the panel.

39. An Improved Method of Tissue Fixation Suitable for Electron Microscopy. CLIFFORD E. GREY, JOHN J. KELSCH, AND MARY C. SCHUSTER, *Interchemical Research Laboratories, New York, New York.*

Standard methods of tissue fixation produce artifacts, while not too objectionable for light microscopy, are too pronounced for electron microscopy. We have developed a technique of tissue fixation which minimizes these artifacts by utilization of fixing agents which fix some cell components without obvious protein precipitation and under others less soluble. The rational is as follows: Immediate immersion of fresh tissue in neutral formalin, which fixes all cellular elements except carbohydrates, fatty substances and chromatin material; a thorough washing in water; immersion in a solution of osmium tetroxide and potassium dichromate which fixes chromatin and fats; without washing, immersion in chromic acid for fixation of carbohydrates and to complete fixation. This method prevents excessive and unequal shrinkage during dehydration and embedding by standard methods. Electron and light micrographs of various tissue prepared by various standard methods and by our improved method are presented illustrating undesirable fixation effects and the delicate fixation of our present technique on such

substances as fats and mucus secretion fixed within a cell sectioned sufficiently thin for electron penetration giving excellent detail of internal structure.

40. Fixation and High Speed Microtomy of Botanical Specimens for the Electron Microscope. J. J. KELSCH, *Interchemical Research Laboratories, New York, New York.*

Portions of the ovary and stem of a specimen of the genus Rosa were prepared for high speed microtomy by the improved fixation method developed and announced by Mr. C. Grey, John J. Kelsch, and Mary C. Schuster (Interchemical Research Laboratories, New York). An extensive study of animal tissue fixation has been done, and the fields of electronmicrohistology and cytology have been opened to the medical profession and the biologist. This paper suggests the use of the high speed microtome and the improved fixation methods in the new field of electromicrobotany. Electron and light micrographs of the Rosa ovary and stem are presented to illustrate the potentialities of the new methods.

41. Some New Bacteriological Techniques in Electron Microscopy. J. HILLIER, *RCA Laboratories, Princeton, New Jersey,* GEORGES KNAYSI, *Department of Bacteriology, Cornell University, Ithaca, New York,* AND R. F. BAKER,* *RCA Laboratories, Princeton, New Jersey.*

It is now recognized that the older methods of mounting bacteria for electron microscope examination in which the organisms are dried on the supporting membrane from a distilled water suspension introduces a number of artifacts. A number of new techniques have been developed, all of which are directed toward subjecting the bacteria to a minimum amount of handling or of changed conditions. The first of these involves coating a culture on a solid medium with a plastic lacquer in such a way that when dried lacquer film is removed from the culture, it carries with it the surface layer. The method is ideal for certain types of growth. It is particularly useful in the study of the plaques produced by the action of bacteriophage. A more generally useful method of preparing the organisms is by growing them directly on a supporting film of collodion placed on the surface of agar. The diffusion of the nutrients through the thin membrane appears to be sufficient for profuse growth. A simple modification of this technique permits it to be applied to organisms which normally grow in liquid media. These techniques are particularly useful for the study of growth cycles in the organism, in the study of spore germination, cell division and the growth of flagella. A number of examples will be shown and some modifications described.

* Present address: School of Medicine, University of Southern California, Los Angeles, California.

42. Nuclei and Other Structures of Bacteria. C. F. ROBINOW, *Strangeways Laboratory, Cambridge, England,* AND V. E. COSSLETT, *Cavendish Laboratory, Cambridge, England.*

Seeking information on the fine-structure of the nuclei and other cytological constituents, which the light microscope reveals in stained preparations of bacteria, we have studied species which are naturally transparent to electron beams accelerated at 50–90 kv, and others in which detail is visible only after special treatment. In osmium- or formalin-fixed cells of a species of *Pseudomonas* the nuclei appeared as characteristically shaped groups of small vesicles and fine granules of much lower density than the surrounding cytoplasm; a result to be expected (re density) from the appearance of the nuclei in the light microscope. In *Bact. carotovorum* and a species of *Corynebacterium*, both larger and denser than the transparent *Pseudomonad*, the nuclei were visible in specimens prepared from young agar plate cultures briefly exposed to glacial acetic acid vapor. Here again the nuclei, identified by characteristic shape and arrangement, proved to be less dense than the cytoplasm. For studying gram-positive spore-forming bacteria a method was used which renders these normally opaque organisms very transparent and permits identification of cell wall, plasma membrane, cytoplasm, polar bodies, and nuclei. The method consists in exposing for a short time young cultures to vapors of ether or ether with 5–10 percent acetic acid or alcohol; the autolytic processes which the ether initiates are then arrested, after a suitable interval, with osmium tetroxide vapor or with 5 percent formalin.

43. Contribution of Electron Microscope Studies to Physiological Studies of Plants. EDITH A. ROBERTS AND MILDRED D. SOUTHWICK, *Department of Plant Science, Vassar College, Poughkeepsie, New York.*

Photosynthesis. Electron micrographs show that chloroplasts throughout the entire plant kingdom are composed of connecting protoplasm units of varying mass and density. The smallest masses are shown to be of less than 100 ängstroms and their interconnecting strands are of correspondingly decreasing dimensions. These same protoplasmic entities are found to be present in "chlorophyll solutions" and "chlorophyll crystals." *Vitamins.* Electron micrographs of carrot root walls show them to be of comparable structure to carotene bodies in cells and that these are composed of protoplasmic entities of less than 100 ängstroms. These are identical in structure with the bodies present in "vitamin A crystals" and suggests, therefore, that vitamin A is present in young, actively growing plant tissues, which may be used as a direct source of vitamin A.

44. Electron Micrographs of *Escherichia Coli* Cells Treated with a Nitrogen Mustard. ABE EISENSTARK, *University of Illinois.*

When *Escherichia Coli* cells are treated with methyl *Bis*-(β-chloroethyl)amine-hydrochloride, bizarre morphological forms result. This peculiar morphology was studied with the electron microscope. Shortly after treatment, the normal rod-shaped bacterial cell begins to show a slight bulging in the center of the cell, accompanied with a clear zone traverse across the cell. After this, the cell elongates and large, clear vacuoles appear at periodic intervals along the cell. Granules appear later inside of these vacuoles.

After about 24 hours, cells no longer seem to be distributed at random but are grouped into small colonies. Under the electron microscope, minute granules surround these small groups of cells. These bizarre forms show some resemblance to the forms obtained when bacterial cells are treated with x-ray.

45. Some Applications of the Delft Electron Microscope in Biology. WOUTERA VAN ITERSON, *National Institute of Health, Bethesda, Maryland.* (Invited paper.)

Though post-war circumstances still impose difficulties on a rapid development of research with the electron microscope in the Netherlands, the Delft instrument constructed by Le Poole has been used for investigations in several divergent fields. As examples, a few micrographs will be shown bearing on (1) existing problems in bacteriology; (2) the structure of chloroplasts; and (3) the fine structure of sperm cells. These electron micrographs have all been made at 90 kv, rather than the 50 kv commonly employed in this country.

46. The Electron Microscopy of Bacteriophage Production. R. W. G. WYCKOFF, *National Institute of Health, Bethesda, Maryland.*

Study of the way viruses originate within the cells they infect promises to give more direct information than is available in any other way about the fundamental nature of these smallest of possible living agents. Bacteriophages are among the best of these virus-like objects for electron microscopic study because of the ease with which detail can be seen within their host bacteria. This paper is a discussion of techniques used for making the necessary preparations, and a comparison of results obtained with various materials as shadowing agents. A series of electron micrographs will also be shown illustrating steps observed in the growth of bacteriophage particles.

47. Observations on the Morphology of *Hemophilus Pertussis*. JULIA M. COFFEY AND SOPHIA M. COHEN, *Division of Laboratories and Research, New York State Department of Health, Albany, New York.*

A correlation was sought between morphology revealed by electron microscopy and other biologic characters of strains of *Hemophilus pertussis*. No significant structural differences were observed between four strains of phase I that varied in antigenicity and virulence for mice. However, morphology changed with age. Cells from young cultures appeared nearly spherical with cytoplasm of relatively uniform density surrounded by an outer zone semitransparent to the beam. In older cultures diffuse and clearly circumscribed areas of greater density occurred in the cytoplasm. In specimens shadowed with gold, marked irregularities in surface were evident, some of which were undoubtedly caused by drying. Certain areas of greater elevation seemed to correspond with those of greater density previously noted. In saline suspensions of *H. pertussis* treated with 1-percent formalin followed by heating one-half hour at 75°C, the semi-transparent zones were ragged and broken, but the cytoplasm remained apparently intact. A similar change was not observed in cells suspended in salt solution with or without formalin in concentrations of 0.1 and 1 percent even after storage at 4°–6°C for three weeks. In these, the outer zones were wide, resembling swollen capsules.

Supplementary Session

48. Focal Properties and Aberrations in Ideal Electron Lens Fields. DR. REINHOLD RÜDENBERG, *Graduate School of Engineering, Harvard University, Pierce Hall, Cambridge 38, Massachusetts.* Only a few field structures permit both rigorous mathematical description and explicit determination of electron trajectories. One of these fields, of geometric symmetry as in electron lenses and derived from the condition of focus formation, is analyzed in detail. Formulations are developed for the various possible electron paths and charts are given for the most outstanding solutions. The focal properties within and without the volume of such a field formation are explicitly derived, and discussed as to the performance of the cardinal planes, the focal length and the back focus, the latter being of great interest with respect to perfect image formation. For thin lenses a simple approximation leads to familiar formulations. For thick lenses the behavior is more complex, but also easy to survey. Expressions are derived for velocity and aperture defects different as to volume and surface aberration, and dependent only on a few parameters of the lens field. Experimental verification of the field structure shows good accord with the predicted electronic properties of the lens field. The systematic results developed may be useful when discussing the properties of actual electron lenses.

49. Recent Electronmicrographs Obtained with a New Electrostatic Electron Microscope. G. FLEMING, R. REMPFER, E. ASHERMAN, AND P. NOLAN, *Farrand Optical Company, Inc., New York City, New York.* An electron microscope is being developed at the Farrand Optical Company. An experimental model has been built which employs an electrostatic lens imaging system. The performance of this model is encouraging as to stability of operation and image quality obtainable. Slides illustrating the results with a variety of specimens will be shown.

50. The C.S.F. Electron Microscope. P. GRIVET, *Compagnie Générale de C.S.F., Paris, France.* The C.S.F. electron microscope is of the purely electrostatic type. Two stages of lenses bring the electronic magnification to a value of 14,000 times, with a resolving power of about 4 millimicrons, for standard specimens and standard practice. The magnification can be adjusted by variation of